

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|--|--|-----------|--|
| (51) International Patent Classification ⁶ : C12Q 1/68 | | A1 | (11) International Publication Number: WO 99/37817 (43) International Publication Date: 29 July 1999 (29.07.99) |
| (21) International Application Number: PCT/US99/01552 (22) International Filing Date: 25 January 1999 (25.01.99) (30) Priority Data: 09/013,496 26 January 1998 (26.01.98) US 09/236,026 22 January 1999 (22.01.99) US (63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application US 09/013,496 (CIP) Filed on 26 January 1998 (26.01.98) (71) Applicant (for all designated States except US): SCHERING AKTIENGESELLSCHAFT [DE/DE]; Muellerstrasse 170-178, D-13342 Berlin (DE). (72) Inventors; and (75) Inventors/Applicants (for US only): JOHNSON, Paul, H. [US/US]; 652 San Benito Avenue, Menlo Park, CA 94025 (US). PONTE, Phyllis, A. [US/US]; 841 Vista Grande, Millbrae, CA 94030 (US). ZAJCHOWSKI, Deborah, A. [US/US]; 2246 15th Avenue, San Francisco, CA 94116 (US). | | | (74) Agents: POWERS, Margaret, A. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111-3834 (US). (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> |
| (54) Title: GENE EXPRESSION METHODS FOR SCREENING COMPOUNDS | | | |
| (57) Abstract A method for screening test compounds <i>in vitro</i> for predicted <i>in vivo</i> activity is disclosed. | | | |

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakhstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

GENE EXPRESSION METHODS FOR SCREENING COMPOUNDS

The present application is a continuation-in-part application of U.S. Patent
5 Application No. 09/013,496, filed January 26, 1998, the disclosure of which is
incorporated herein by reference in its entirety for all purposes.

BACKGROUND

Differences in the expression of genes in normal versus activated,
10 diseased, neoplastic cells or the like can be helpful in understanding cellular processes
resulting in the affected state. For example, Zhang *et al.* (Science 276:1268-1272
(1997)) disclosed gene expression patterns in gastrointestinal tumors, identifying more
than 500 transcripts that were expressed at significantly different levels in normal and
neoplastic cells. Bernard *et al.* (Nucl. Acids Res. 24:1435-1442 (1996)) disclosed a
15 method for analyzing the expression levels of 47 genes in resting and activated T cells,
as well as in epithelial cells.

Microarrays of synthetic oligonucleotides or cDNAs are useful in
evaluating differential gene expression. For example, Schena *et al.* (Science 270: 467-
470 (1995)) disclosed the quantitative monitoring of gene expression patterns in response
20 to transgenes using a complementary DNA microarray. Shena *et al.* (Proc. Natl. Acad.
Sci. U.S.A. 93(20):10614-10619 (1996)) used microassays containing human cDNAs of
unknown sequence to quantitatively monitor differential gene expression patterns under
given experimental conditions. De Risi *et al.* (Nat. Genet. 14(4):457-460(1996)) used a
cDNA microarray to analyze gene expression patterns in human cancer. Heller *et al.*
25 (Proc. Natl. Acad. Sci. U.S.A. 94(6):2150-2155 (1997)) disclosed the use of cDNA
microarray technology to monitor gene expression in inflammation.

Other methods for screening include a method for detecting and isolating
differentially expressed mRNAs using first oligonucleotide primers for reverse
transcription of mRNAs and both the first oligonucleotide primers and second
30 oligonucleotide primers for amplification of the resultant cDNAs (U.S. 5,580,726).
Rosenberg *et al.* (PCT Publication WO 95/21944) disclosed the use of expressed
sequence tags (EST's) to detect genes differentially expressed in healthy subjects vs.
subjects having a disease of interest. Lee *et al.* (Cell Biology 92:8303-8307 (1995))

disclosed the use of comparative expressed -sequence -tag analysis to identify about 600 differentially expressed in RNAs in untreated and nerve growth factor-treated PC12 cells.

Further screening methods include such examples as that of Nilsson *et al.* (PCT Publication WO 93/07290) who disclosed an *in vitro* method of evaluating the antagonistic vs agonistic effects of a receptor-binding substance on selected types of cells containing endogenous intracellular hormone receptors by analyzing cellular response to the receptor-binding substance based on the level of expression of the protein product made by a gene regulated by the hormone-receptor interaction. WO 96/41013 disclosed a method for identifying a receptor agonist or antagonist using mutant versions of intracellular receptors such as the estrogen (ER), androgen (AR), progesterone (PR), and glucocorticoid (GR) receptors.

Knowledge that environmental agents alter gene expression has led to the employment of specific genes as biomarkers of exposure to chemicals and other environmental factors (Links *et al.* (Annu. Rev. Public Health 16:83-103 (1995))). Such biomarkers have been used to screen chemicals and biological samples for ability to alter gene expression (Sewall *et al.* Clin. Chem. 41:1829-1834 (1995)).

Thus, a need exists for methods to screen and characterize differential gene expression *in vitro* and to screen compounds for their effects on gene expression *in vitro*. The instant invention addresses these needs and more.

SUMMARY OF THE INVENTION

One aspect of the invention is a method for grouping test compounds into classes, the method comprising:

- (a) exposing a cell culture or cultures comprising at least two gene-cell combinations to a test compound to generate an exposed cell culture or cultures;
- (b) preparing RNA from the exposed cell culture(s);
- (c) screening RNA from (b) for mRNA of each gene in the gene-cell combinations of (a) to generate a gene expression fingerprint (GEF) for the test compound;
- (d) repeating steps (a) - (c) for each test compound to be grouped in classes; and
- (e) comparing the GEF for each test compound (d), wherein the test compounds are grouped into at least two classes based on differences in their GEFs.

Representative test compounds in each class may be further tested for a representative activity or an activity of interest *in vivo*.

The at least two gene-cell combinations may, for example, comprise at least two different genes, at least two different cell types, or combinations thereof. In some embodiments a gene or genes in the gene-cell combinations may comprise an endogenous gene under control of its native promoter, a heterologous gene under control of a heterologous promoter, an internal negative control gene, wherein an effect on the mRNA level of the negative control gene in response to the test compound is indicative of a toxic effect of the test compound, or an internal negative control gene, wherein the effect on the mRNA level of the negative control gene in response to the test compound is indicative of a non-specific effect of the test compound.

Screening of the RNA may comprise PCR amplification using oligonucleotide primers specific for each gene. In some embodiments, the RNA is optionally reverse transcribed into cDNA. In some embodiments, the screening comprises hybridization of nucleic acid sequences specific for each gene to the RNA or cDNA of the exposed cell cultures. In further embodiments, the level of the mRNA of at least one gene in the at least two gene-cell combinations is quantitated.

In some embodiments of the invention, combinations of two or more test compounds can be administered to the cell cultures to generate a GEF for the combination.

A further aspect of the invention is a method of identifying one or more genes for use in a gene-cell combination for grouping test compounds into classes, the method comprising:

- (a) exposing host cells *in vivo* or at least one host cell culture to a first reference compound;
- (b) preparing RNA from the host cells *in vivo* or host cell culture of (a); and
- (c) comparing the RNA of (b) to RNA from host cells *in vivo* or a control host cell culture not exposed to the first reference compound; wherein at least one gene having an mRNA level affected in response to the first reference compound is identified as a gene for use in a gene-cell combination for grouping test compounds into classes. The RNA of (c) may be compared to RNA from host cells *in vivo* or a control host cell culture, wherein the host cells *in vivo* or a control host cell culture have or has been exposed to a second reference compound, whereby a gene having an mRNA level

affected in response to the first reference compound but not the second reference compound is identified as having a response specific for the first reference compound.

A further aspect of the invention is a method for grouping test compounds into classes, the method comprising:

- 5 (a) exposing a cell culture or cell cultures comprising at least two gene-cell combinations to a test compound to generate exposed cell cultures, wherein at least one gene in the at least two gene-cell combinations is differentially expressed in a first and second reference state, to generate exposed cell cultures;
 - (b) preparing RNA from the exposed cell culture or cultures;
 - 10 (c) screening RNA from (b) for mRNA levels of each gene in the gene-cell combinations of (a) to generate a gene expression fingerprint (GEF) for the test compound;
 - (d) repeating steps (a) - (c) for each test compound to be grouped into classes; and
 - 15 (e) comparing the GEF for each compound tested in (d);
- wherein compounds are grouped into at least two classes based on differences in their GEFs. In some embodiments at least one of the first and second reference states is a disease state such as cancer.

In another aspect, the invention provides a method of generating a reference gene expression fingerprint (GEF) for at least one reference compound for use in grouping test compounds into classes, said method comprising:

- (a) identifying at least two gene-cell combinations, each of said at least two gene-cell combinations comprising a unique combination of a particular gene and a cell of a particular cell type, wherein a first gene-cell combination is identified by:
 - 25 (i) exposing host cells in vivo or a host cell culture of a first cell type to a first reference compound;
 - (ii) preparing RNA from the exposed host cells in vivo or the host cell culture of (ii);
 - (iii) comparing the RNA of (ii) to RNA prepared from host
- 30 cells in vivo or a host cell culture of the first cell type not exposed to the first reference compound, wherein a change in a level of mRNA for a gene in cells of the first cell type in response to the first reference compound identifies the gene and cells of the first cell type as the first gene-cell combination for grouping test compounds into classes; and wherein a second gene-cell combination is identified by:

(iv) exposing host cells in vivo or a host cell culture of the first cell type or a second cell type to the first reference compound;

(v) preparing RNA from the exposed host cells in vivo or the host cell culture of (iv);

5 (vi) comparing the RNA of (v) to RNA prepared from host cells in vivo or a host cell culture of the same cell type as in (iv) not exposed to the first reference compound, wherein a gene having an mRNA level changed in response to the first reference compound is identified as a gene for use in the second gene-cell combination for grouping test compounds into classes, said second gene-cell combination
10 being different from said first gene-cell combination and comprising the identified gene and cells of the same cell type as in (iv); and

(b) screening RNA of (ii) and (vi) for mRNA for each gene in each of the at least two gene-cell combinations to generate a reference GEF for the first reference compound for use in grouping test compounds into classes.

15 In another aspect, the invention provides a method for grouping test compounds into classes, said method comprising:

(a) generating a reference GEF for a reference compound according to the method described immediately above and discussed below;

(b) generating a GEF for each test compound to be grouped into
20 classes by:

(i) exposing a cell culture or cultures comprising the at least two gene-cell combinations identified in claim 1 to a test compound to generate an exposed cell culture or cultures;

(ii) preparing RNA from the exposed cell culture or
25 cultures of (i);

(iii) screening RNA of (ii) for mRNA of each gene in each of the at least two gene-cell combinations of (i) to generate a GEF for the test compound;

(iv) repeating (i) - (iii) for each test compound to be
30 grouped in classes to generate a GEF for each said test compound; and

(c) comparing the GEF for each test compound generated in (b) with the reference GEF of (a), wherein the test compounds are grouped into at least two classes based on differences or similarities between their GEFs and the reference GEF.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 comprises Figures 1A and 1B. Figure 1A is a graphical depiction of GEF results for a reference compound (Ref) and test compounds x, y, z in two assays. Figure 1B depicts GEF results for a Reference (Ref) compound and seven test compounds in three assays. Each of the squares represents the results of one assay. Activity of a compound in a particular assay is indicated by a solid square. Inactive compounds are indicated by an open square.

Figure 2 comprises Figures 2A and 2B. Figure 2A depicts GEF results for a Reference (Ref) compound and six test compounds in five assays. Figure 2B is a single linkage tree diagram showing the percent disagreement between the reference and six test compounds with the GEF activity results depicted in Figure 2A.

Figure 3 comprises Figures 3A-3C. Figure 3A shows consensus GEFs for human breast cells from normal and different stages in malignant progression.

Consensus gene expression changes representative of all of the cell lines classified as either weakly or highly invasive are graphically depicted. The values correspond to the median fold-change relative to the MCF10A reference observed for each gene from data in Tables 7A-7B. The data shown for the "normal" GEF are changes in gene expression observed in the 76N MEC strain relative to MCF10A. Genes with expression changes that are "tumor-associated" are represented by bars with left-handed stripes (bars having a stripe angling downward from left to right), genes associated with weakly invasive cancers have solid bars, and genes associated with highly invasive cancers with right-handed stripes (bars having a stripe angling upward from left to right). The stippled bars denote genes whose direction or extent of expression change is associated with either weakly or highly invasive cancers. The figure legend to the right of the three graphs lists the genes depicted. Each number on the legend identifies a particular gene.

Figure 3B shows GEFs of two breast cell lines with unknown invasive activity. Changes in gene expression of the breast fibroadenoma cell line 006FA2B and the breast epithelial cell line HBL100 relative to MCF10A were determined using Atlas I cDNA hybridization arrays. Data are shown for the 28 genes shown in the figure legend in Figure 3A. The graphical representation of a particular bar (left-handed stripe, right-handed strip, stippled, or solid) has the same meaning as set forth above for Figure 3A.

Figure 3C depicts GEFs for tumor biopsy specimens. Gene expression was monitored by analysis of tumor RNA using Atlas I cDNA hybridization arrays. Changes in gene expression relative to a normal breast tissue specimen for the 28 genes

listed in the figure legend of Figure 3A are shown. The graphical representation of a particular bar (left-handed stripe, right-handed strip, stippled, or solid) has the same meaning as set forth above for Figure 3A.

Figure 4 shows gene expression changes following treatment of MDA231 with various compounds. MDA231 cells were exposed to taxol, butyrate, mevastatin, or vehicle control for 72 h and analysed for effects on gene expression as described in M&M. The data shown correspond to effects on mRNA levels elicited by drug treatment relative to control for those genes that had greater than 2-fold changes in expression in at least one treatment condition.

DETAILED DESCRIPTION OF THE INVENTION

I. Overview of Methods

The instant invention is directed to screening methods that allow the grouping of compounds into classes of compounds with similar activity(s), as measured by the changes elicited by the compounds in the expression of certain genes in certain cells. There is no requirement that the certain genes or cells employed in the analysis be identified by function, map location, or other parameter physiologically relevant to a disease or indication for which a therapeutic drug is intended or sought.

Typically, a reference "gene expression fingerprint" (GEF) is first generated for a reference compound or "state". A GEF is then generated for each test compound of interest as a result of the screening process of the invention. The test compounds are then grouped into classes on the basis of comparison with the reference GEF.

The basic screening process used herein to generate the reference GEF or to screen test compounds relies on the use of "gene-cell combinations". A "gene-cell combination" as used herein refers to a particular gene in a particular host cell type. Different gene-cell combinations can arise from various combinations of particular genes and particular host cell types, such as the same gene in two or more host cell types, two or more different genes in the same host cell type, and so on. In addition, a single host cell may comprise one or more such genes to generate two or more gene-cell combinations.

A host cell type as used herein refers to a cell of a particular source, such as but not limited to tissue of origin, state of differentiation, adaptation to particular

growth conditions, clonal variants, cell line, transformation, transduction, viral infection, parasite infection, bacterial infection, transgenic host, species of origin, and so on.

Thus, for example, in an embodiment a reference GEF is generated for a reference compound by exposing a cell culture or cultures comprising at least two gene-cell combinations to the reference compound and observing a change in the mRNA level(s) of the gene(s) in the gene-cell combinations in response to the reference compound. In a preferred embodiment of the invention, a single gene-cell combination is considered insufficient to generate a GEF. More typically, several gene-cell combinations (also termed herein "assays") are examined in response to the reference compound or in comparison of reference states to generate a "reference GEF".

In yet a further embodiment of the invention, the relative mRNA levels of at least one gene are compared in at least two host cell sources, wherein each host cell source comprises a different reference state to generate a reference GEF for a reference state. As discussed herein, the genes are chosen on the basis of being differentially expressed in a first and second reference state. Typically, at least one of the reference states is a disease state.

In the screening of test compounds by the methods of the invention, test compounds or agents, such as libraries of peptides, peptidomimetics (such as, but not limited to p53, estrogen, raloxifene, tamoxifen, or IFN β mimetics), polypeptides, proteins, ribozymes, nucleic acids, oligonucleotides, or other organic or inorganic compounds, or natural products (*e.g.*, microbial broths, plant or animal cell extracts) are subjected to a screening process in which a GEF is generated for each test compound by exposing a cell culture or cultures comprising at least two gene-cell combinations to each compound and observing any changes in the mRNA level(s) of the gene(s) in the gene-cell combinations in response to the test compound. The results are used to compare similarities and differences among the test compounds screened. Based on these similarities or differences, the test compounds are divided into groups for further analysis. Such further analysis may involve *in vivo* testing or further screening in other assays.

In some embodiments of the invention, the methods of the invention are useful to identify compounds or agents that, for example, are mimetics of protein function (*e.g.* p53-induced changes in gene expression) or modulate a disease-associated GEF in the direction of an unaffected GEF (*e.g.*, neoplastic vs. "normal", atherosclerotic plaque vs. "normal" blood vessel, inflammatory tissue vs. "normal" tissue). In such

cases, the "reference GEF" is preferably derived from the differential gene expression patterns observed between different cell states (*e.g.*, p53 positive vs. negative; metastatic vs. non-malignant tumors) and not necessarily from treatment with a reference compound per se.

5

II. Reference Compounds and States

As used herein, the reference compound may comprise a protein, polypeptide, peptide, nucleic acid, peptidomimetic, ribozyme, nucleic acid, oligonucleotide, or other organic or inorganic compound, or microbial, plant, and animal natural products. The reference compound is preferably chosen as having a representative *in vivo* activity, such as, but not limited to, inhibition of cell growth, stimulation of a receptor of interest, catalysis of a compound of interest, synthesis of a compound of interest, inhibition of replication of a virus of interest, stimulation of cell growth, inhibition of cell invasion of extracellular matrix, chemotactic response, anti-metastatic activity, anti-atherosclerotic activity, anti-inflammatory activity, anti-apoptotic effects, prevention of atherosclerotic lesion progression, decreased bone loss, decreased inflammation in rheumatoid arthritis, improved cognitive function, or prevention of hot flushes. However, the GEF generated for the reference compound need not directly be a measure of such activity. Rather, the GEF need only be representative of the effect on mRNA levels of the reference compound in a given gene-cell combination, or set of gene-cell combinations. Furthermore, the genes assayed for mRNA levels need not be directly or indirectly involved with the desired *in vivo* activity. In the screening methods of the invention, test compounds are screened to allow grouping into classes relative to the reference compound. Members of such classes can then be screened for the desired *in vivo* activity, lack of side effects, or other improved features.

One of ordinary skill in the art will typically understand that a reference compound is chosen on the basis of the problem to be addressed. Thus, in general, to practice the methods of the invention a reference drug, chemical compound, protein, peptide, oligonucleotide, *etc.* that has a known or predictable physiological effect relevant to a pathological state or desired pharmacologic property is selected as a basis for identification of a class of compounds.

Some exemplary reference compounds include but are not limited to tamoxifen, raloxifene, interferon α (IFN α), interferon β (IFN β), interferon γ (IFN γ), or an anti-Ha-ras-ribozyme (Kijima *et al.*, Pharmacol. Ther. 68:247-267 (1995)); ligands

for nuclear receptors that are transcription factors, such as steroid hormones, retinoids, *etc.*; receptors such as endothelin; ligands for transmembrane receptors, such as endothelin, gastrin releasing peptide, neuregulin, PDGF, cytokines, chemokines, and insulin; extracellular matrix components such as vitronectin, laminin, and collagen; cell
5 adhesion molecules such as N-CAM or I-CAM; inhibitors or activators of an enzyme of interest, such as L-NAME for nitric oxide synthase; chemotherapeutic agents, such as cisplatin or taxol.

A reference compound can also be the product of a gene expressed within a host cell. Such genes may be endogenous or heterologous, under the control of an
10 endogenous or heterologous promoter, *etc.* Exemplary genes include, but are not limited to transgenes, viral genes, antisense nucleic acids, ribozymes, *etc.*

In some cases, a reference state will be employed instead of, or in conjunction with, a reference compound for the determination of the reference GEF. The differences in mRNA levels between two or more cells or tissues representing
15 relevant physiological/pathological states form the basis of a reference GEF. Some examples of reference states include, but are not limited to, normal vs. atherosclerotic blood vessels of varying lesion severity; normal vs. progressive stages in the development of malignant carcinomas, sarcomas, melanomas, or lymphomas; normal vs. stages of neurodegeneration associated with different types and severity of Multiple
20 Sclerosis, Alzheimer's or Parkinson's disease.

III. Gene-Cell Combinations

A. Genes

The instant invention utilizes changes in the mRNA levels of one or more
25 genes in at least two gene-cell combinations, wherein the mRNA level of the gene(s) is responsive to the reference compound, to generate a GEF for each test compound screened. The test compounds may affect mRNA levels directly or indirectly, by, for example, binding to a promoter or other regulatory element, binding to a receptor and triggering some intracellular signal, altering the stability of the mRNA, binding to an
30 intracellular enzyme, such as a kinase or phosphatase, binding to a transcription factor, altering the redox environment, or affecting ion flux into and within the cell. The genes are preferably endogenous genes under the control of their native promoters. In some embodiments, cells may be infected with viruses, wherein the responsive genes are viral genes. In some embodiments, a marker gene, such as a heterologous gene under control

of a heterologous promoter, is introduced into the cell as an internal control for monitoring gene expression or the physiological state of the cell.

The set of one or more responsive genes for screening may be determined in many ways. For example, the mRNA from a cell culture exposed to a reference compound can be compared to mRNA from a control, or unexposed cell culture. In some embodiments, an organism or animal is exposed to a reference compound *in vivo*, and the organism, tissue samples, explants, primary cultures, or the like used as the source for mRNA. Changes in the level of specific mRNA that occur in response to the reference compound can be identified by a variety of means, including but not limited to subtractive hybridization using either normalized or unnormalized libraries (*e.g.*, Gurskaya *et al.*, Anal. Biochem. 240:90-97 (1996), Bonaldo *et al.*, Genome Res. 6:791-806 (1996)), the use of multiple arrays made with EST's or cDNAs (*e.g.*, Bernard *et al.*, Nucl. Acids Res. 24:1435-1442 (1996); Schena *et al.*, Science 270:467 (1995)), DD-PCR (Liang *et al.*, Science 257:967-971 (1992)), SAGE (Velculescu *et al.*, Science 270:484 (1995)), *etc.*

Although it is not required for the instant invention that the responsive genes be responsible for any desired *in vivo* effect of the reference compound, it may be advantageous to use responsive genes of known identity and function. For example, genes known to be responsive to the reference compound may comprise all or part of the set of responsive genes. Such genes may be identified from the literature, from cloning of cDNAs from cell cultures exposed to the reference compound, or other source. Thus, for example, epidermal growth factor-regulated genes such as junB, rhoB, EGF receptor, integrin beta 1, and viculin may comprise all or a part of a set of genes to screen candidate compounds for selective EGF receptor agonists or antagonists. Genes encoding such proteins as p21, MDR1, hsp70, IGFBP-3, and bax have all been shown to be regulated by p53 through different mechanisms. These genes may comprise all or a part of a set of genes to screen candidate compounds for p53 mimetics.

Preferably, a responsive gene chosen for use in the screening assay sustains at least a two to fivefold change in the level of its mRNA in response to the reference compound. This change may be an increase or decrease. The measure of five-fold or greater responsiveness provides for the detection of "weakly " active test compounds which may, for example, provide only a "partial" response (*e.g.*, a two-fold change in mRNA levels in comparison with a "full" response that is five-fold).

In some embodiments of the invention, the same set of responsive genes, or a subset thereof, or yet a different set, is examined in more than one cell type as part of the screening (*i.e.*, to generate different "gene-cell combinations").

Preferably two to 15 or more gene-cell combinations (or "assays") are used in screening compounds. The number of assays used to characterize compounds or reference states into groups based on GEF can be reduced using additional reference compounds with known *in vivo* effects. GEF's can be interpreted as like or unlike the reference compound or state. For example, when the additional reference compound has undesirable *in vivo* effects, assays which fail to distinguish the additional reference compound from the first reference compound may be eliminated from the screening used to generate GEFs. Some of the gene-cell combinations may be internal controls. For example, "house-keeping" genes such as GAPDH, actin, or cyclophilin are typically expected not to respond to the reference compound and thus can serve as negative internal controls. Positive internal controls can comprise, for example, a recombinant molecule under control of a promoter expected or known to be responsive to the reference compound.

Additional internal controls can comprise genes which are predictive of possible "toxic" effects of the reference or test compounds. For example, such control responsive genes include but are not limited to cytokines such as TNF or lymphotoxin, heat shock proteins such as hsp70, DNA damage inducible genes such as gadd153 or gadd45, and the like. An increase in the mRNA level of one or more of these genes is typically predictive of a toxic effect of the reference or test compound. Thus, for example, in an embodiment, screening of test compounds for reduced toxic effects is accomplished by looking for reduced or unchanged levels of these internal control genes.

B. Cells

Typically, a cell line and gene are chosen in concert as an "informative" gene-cell combination for the screening of test compounds. Practical considerations include the tissue of origin of the cell line; the level of differentiation of the cell line, the level of expression of the target genes, the efficiency with which compounds such as cDNA, peptides, ribozymes, and so on can be taken up by the cell line, and so on. In some embodiments, tissue explants or clinical samples such as primary cell cultures, tissue explants from experimental animals, or clinical specimens such as blood samples, tumor biopsies, atherosclerotic blood vessels from a patient are preferred. Thus, for

example, although not a requirement in the instant application, it may be advantageous in the screening of compounds wherein the goal is to develop a new prostate tumor therapeutic to use a prostate cell line.

5 IV. Screening methods

Typically, test compounds, preferably in the form of a library, are screened against the set of responsive genes and cells to identify the compounds with identical or similar gene expression patterns. In an embodiment, for example, a library of about 10^5 - 10^7 test compounds (*e.g.*, peptides, oligonucleotides, ribozymes, peptidomimetics, polypeptides, proteins, nucleic acids, oligonucleotides, or other organic or inorganic compounds, *etc.*) is screened. For example, a small molecule library is screened by exposing cell cultures to a typical final concentration of test compound of 1 - 10 μ M. A range of concentrations (*e.g.*, low, medium, high) for each test compound is preferred to enable the detection of weakly active compounds and to help distinguish compounds which have different levels of activities at given concentrations. For convenience, the cell culture treatment may be in 96 well microtiter dishes. Exposure is typically done for a period of 24 to 48 hours, but can be as short as 30 minutes or as long as a week, especially in the case of transfected or infected cells. The cells are usually treated in a humidified environment containing 5 to 10% CO₂ at 37°C, but variations on these conditions may be warranted by the specific screen. RNA is then recovered from the exposed cultures by methods well known in the art, preferably by a method readily adapted to high throughput (*e.g.*, 96 well format) such as, but not limited to, poly dT capture plates (Mitsubishi *et al.*, Nature 357:519-520 (1992)) or silica gel-based membrane adsorption purification (*e.g.*, Qiagen's RNeasy Total RNA Extraction Kit). The mRNA may be optionally reverse-transcribed into cDNA. The mRNA or cDNA can be used as probe or as target in hybridization reactions, and may be immobilized or in solution. Messenger RNA from the set of one to twenty or more responsive genes can be quantitated by methods well known in the art using such exemplary techniques as standard Northern or slot blot hybridization, nuclease protection, or quantitative PCR which are limited in the number of different RNAs that can be simultaneously analyzed as well as in their amenability to automation. Other preferred methodologies employ isotopically or fluorescently-labeled RNA or cDNA prepared from the isolated cellular RNA as hybridization probes for arrays containing purified cDNAs spotted onto membrane filters (*e.g.*, Bernard *et al.*, Nucl. Acids Res.

24:1435-1442 (1996)) or glass slides (Skena *et al.*, Science 270:467-470 (1995)). A modification of this general methodology utilizes chemically synthesized oligonucleotides covalently attached to a solid substrate instead of cDNAs as the target of the hybridizing RNA or DNA (Lockhardt *et al.*, Nature Biotech. 14:1675-1680 (1996)). An alternative method directly measures the RNA or cDNA by hybridization with gene-specific oligonucleotides, that can be differently labeled (*e.g.*, with mass labels that can be quantitated by time-of-flight (TOF) mass spectrometry; fluorescence enhancers, such as europium, terbium, samarium, and dysprosium, and the like (Xu *et al.*, Anal. Chem. Acta. 256:9-16 (1992)).

10 The GEF for each compound comprises the results of the screening procedures. Compounds may be eliminated from further testing because of the likelihood of toxic effects on the cell, nonspecific responses elicited, and so on. The GEF may be further modified by further testing with additional responsive gene - cell combinations, by using the same set of responsive genes and cells but different concentrations of test compounds, eliminating uninformative responsive gene-cell combinations from the GEF, and so on.

VI. Grouping Test Compounds into Classes

20 Test compounds screened as discussed above are then sorted into classes based on their GEFs. For example, test compounds which elicited a change in mRNA levels of all members of a set of responsive gene-cell combination would be grouped separately from test compounds which elicited a change in only one instance, two instances, *etc.* As the number of assays used for screening increases, more grouping becomes possible.

25 Thus, for example, the reference compound is defined as being "active" in all GEF assays; activity can be an increase or decrease, relative to control, in the mRNA level for the particular gene following compound treatment. A compound x or y is discovered or identified by having activity in at least one GEF assay. Compounds x and y are categorized separately from the reference compound based upon inactivity in at least one assay.

30 Compounds are categorized with each other if they are active in the same assays. In the simplest example employing two assays (see Figure 1A), four possible categories of compound can be defined. The number of possible categories is equal to x^n , where x is the number of activity states measured (*e.g.* + and -) and n is the number

of assays. In this example $x^n = 2^2$ or 4 possibilities, represented by the reference and compounds x, y, z. Each compound is distinguishable from the others by a different GEF. The categories can be further refined by considering quantitative differences in the response to different compounds as a criterion for classification.

5 By increasing the number of GEF assays that are evaluated, more categories of compounds can be defined. Compounds that are active in the same assays are categorized together. In the example in Figure 1B, where $x^n = 2^3$ or 8 possibilities; the seven compounds (x, y, z, a, b, c, d) are representative of different categories. In situations where there are three or more assays (*e.g.*, Figures 1B, 2A, and 2B),
10 clustering algorithms can be used to determine the similarity of each compound to the reference compound and to each other. Initially, compound categories can be determined by their linkage distance, which is a measure of the percent of disagreement with the reference. When a compound shows a high percentage of activity matches with the reference, the closer the linkage distance is between a compound and the reference. By a
15 simple clustering algorithm based on similarity to the reference, the compounds shown in Figure 2A would be characterized by the linkage diagram in Figure 2B. In this analysis, compound z is closest to the reference (*i.e.* linkage distance of 0.4) and compounds a and x are at equivalent distance. By changing the criterium for categorization to a linkage distance of 0.6, both of these compounds could be categorized with z. Thus, the
20 stringency of the categorization can be adjusted by changing this linkage distance. Use of smaller linkage distances as the criteria for categorization would result in the generation of more categories than those obtained using greater linkage distances. Depending upon the data set, additional algorithms can be used to cluster the compounds based upon similarity to each other (James, M., Classification Algorithms (1st ed.) New
25 York, NY, John Wiley & Sons (1985)).

The compounds with activity in only one assay (or less than 20% of the assays, when there are greater than eight assays) are not categorized or further evaluated unless they are active in assays that form the basis for the majority of the active
30 compounds identified (indicating that they may be affecting a portion of the same signaling pathway). For example, in Figure 2A compounds y and b would be potential candidates for further evaluation because they are active in assays that identify compounds x, a, and z. Compound c would not be further tested.

The decision to increase the stringency for categorization can be influenced by the pattern of gene expression observed as well as data from other assays. For

example, in Figure 2A if evaluation of compounds x, z, and a revealed that only x and z were active in an important cell-based assay, compounds such as b and y which demonstrate activity in assays common to x and z would be further evaluated alone and in combination.

5

VII. Further Evaluation of Test Compounds

After grouping of the test compounds into classes on the basis of GEF, representative compounds can be further characterized in cell-based assays well known in the art for properties of interest. Such assays might include, for example, inhibiting or stimulating effects on cell growth, anti-viral activity, gel electrophoretic mobility shift assays with DNA-protein complexes prepared from extracts of treated cells, cell invasion through extracellular matrix or reconstituted basement membrane, anchorage-independent growth, chemotaxis, apoptosis, differentiation, cell adhesion to various substrata, cell-cell interactions, secretion, proteolytic activity, osteoclastic bone resorption, *etc.*

It is advantageous in some instances to extend the cell-based assay to animal models where available. Some examples of animal models known in the art include animal models for uterotrophic effects (*e.g.*, uterine hypertrophy; Allen-Doisey), fever (*e.g.*, rabbit pyrogenicity), osteoporosis (*e.g.*, rat cortical and trabecular bone density following ovariectomy or transgenic/knock-out animals), atherosclerosis (*e.g.*, lipid deposition in blood vessels of rabbits fed lipid-rich diets or in transgenic/knock-out animals), restenosis (*e.g.*, neo-intimal thickening following carotid injury), cancer (*e.g.*, tumor induction in rats or mice, tumor xenograft growth in nude, athymic or in transgenic/knock-out mice), metastasis (*e.g.*, lung colonization following tail vein injection of tumor cells), rheumatoid arthritis (*e.g.*, adjuvant-induced joint swelling), multiple sclerosis (*e.g.*, EAE model in marmosets or rats, transgenic/knock-out mice), Alzheimers disease (*e.g.*, transgenic/knock-out mice).

In some embodiments, the GEF's of two or more test compounds may complement each other, *i.e.*, when the GEF's are superimposed they approximate that of the reference compound or desired aspects of the GEF of the reference compound. In those instances the two or more test compounds may be used together in combination in cell-based or *in vivo* assays to determine whether the combination has desired bioactivity.

The following examples are included for illustrative purposes and should not be considered to limit the present invention.

EXPERIMENTAL EXAMPLES

I. Selective Estrogen Compound Discovery

A. Background

Epidemiological and experimental data support a protective role for estrogen in reducing the incidence and severity of coronary artery disease, Alzheimer's disease, and osteoporosis. Estrogen treatment can, however, lead to unwanted effects such as endometrial hyperplasia in women and reduced testosterone levels in men. Therefore, the aim of the studies described here was to determine whether an *in vitro* profile for compounds with selective *in vivo* protective effects on bone (*e.g.*, reducing bone loss), neuronal function (*e.g.*, anti-Alzheimer's disease), and the vascular system, (*e.g.*, anti-atherosclerotic) could be identified. Such selective compounds would preferably be devoid of undesirable side effects (*e.g.*, uterotrophic effects in females; testosterone-lowering and decreased sex organ weight in males).

The research strategy we have pursued relies on three basic assumptions: these "estrogenic" biological effects are mediated, at least in part, by the estrogen receptor (ER), which is a ligand-inducible transcription factor (Mangelsdorf *et al.*, Cell 83:835-839 (1995)), regulation of gene expression by estrogen occurs by a limited number of mechanistically different processes that may be further modified in a tissue-specific manner, and compounds that have selective *in vivo* effects will elicit distinguishable gene expression patterns.

Available methods for identifying ER ligands that have potential as selective drugs *in vivo* include standard ER ligand binding and cell-based estrogen (E)-dependent proliferation assays, or ER-mediated transactivation assays (*e.g.*, Tzukerman *et al.*, Mol. Endo. 8:21-30 (1994)), which utilize different E-responsive promoters to characterize compounds. Screening for ligands that differ in their abilities to change ER conformation is possible using a proteolytic fragmentation assay (Beekman *et al.*, Mol. Endo. 7:1266-1274 (1993)). Prudent use of these assays can permit the separation of E agonists from partial agonists and antagonists. However, these methods do not provide sufficient information about a compound to enable prediction of *in vivo* selectivity since compounds with markedly different *in vivo* effects are not distinguishable by those assays.

A method to classify compounds based upon differential gene expression modulation was developed herein to identify such selective compounds. A total of forty-nine compounds was tested by this method and thereby categorized into classes

based upon their GEFs. Finally, the *in vivo* activities of some of the sorted compounds were evaluated to determine the predictability of the *in vitro* "fingerprint" for *in vivo* effects.

5 B. Specific Strategy

1. Genes and Cells

Known E-responsive genes were identified by literature search (52kD cathepsin D, growth hormone, prolactin, progesterone receptor, pS2, TGFalpha, IGFBP-1, CBG, Amphiregulin, TRHR (thyroid releasing hormone receptor)) and the
10 corresponding cDNA (or fragments thereof) were cloned and probe fragments prepared for Northern or slot blot hybridization studies by techniques known in the art. Mammalian cell lines that contain endogenous ER were identified through literature reports (GH3 pituitary adenoma, BG-1 ovarian carcinoma, MCF7 breast carcinoma, ZR75-1 breast carcinoma, MDA361 breast carcinoma, Ishikawa human endometrial
15 carcinoma (Nishida *et al.*, Acta Obstet. Gynaec. Jpn. 37:1103-1111 (1985))) and/or by analysis for ER expression (*e.g.*, protein by Western blot analysis; RNA by RT-PCR). In addition, transfected cells which stably express ER were also tested (MDA231-ER--breast carcinoma (Zajchowski *et al.*, Cancer Res. 53:5004-5011 (1993)), 185B5-ER--human mammary epithelial cell line (Zajchowski *et al.*, Mol.
20 Endocrin. 5:1613-1623 (1991)), HepG2-ER--human hepatocellular carcinoma, and Fe33--rat hepatoma (Kaling *et al.*, Mol. Cell. Endo. 69:167-178 (1990))).

The first step was to determine which of the genes and cell lines actually showed measurable responses to E treatment. To that end, ER-positive cells were grown in estrogen-free culture medium and treated with the natural hormone, 17 β -estradiol
25 (E2), or 17 α -ethinyl-estradiol (EE; non-metabolizable estrogen) for short (3h), intermediate (24h), and long (72h) time periods and RNA prepared from the cells at each time point. Analysis of the levels of mRNA for the genes of interest gave an estimate of the kinetics of the response to EE treatment and an indication of the optimal conditions to measure the responsiveness of each gene.

30

2. Grouping of active, specific compounds according to GEF;
 selection of "informative" assays

At this stage, all of the identified E-responsive gene-cell combinations could have been employed in a screen of a large number of compounds. However, for

this concept validation experimentation, we decided to simplify the GEF screen by asking whether a subset of these gene-cell combinations would be sufficient to identify known pharmacologically different compounds. To do this we chose to test those gene/cell combinations that responded to E2 or EE treatment with at least three-fold effects on mRNA level with seven additional compounds. The seven other compounds were chosen based upon known properties in *in vitro* and *in vivo* assays. Important compounds were tamoxifen (the 4-OH-tamoxifen (HT) derivative was used in the initial studies) and raloxifene (Ral) because at the time these studies were carried out, no reported *in vitro* method distinguished them even though they were clearly different in their *in vivo* responses (*e.g.*, although they have comparable anti-estrogenic effects on the mammary gland, tamoxifen is significantly more uterotrophic than is raloxifene (Sate *et al.*, FASEB J. 10:905-912 (1996)). These compounds therefore became additional reference compounds in the analysis, since we wanted to find compounds similar to them as well as different ones. We also chose a compound structurally related to estradiol (*i.e.*, 2-OH-17 β -estradiol (2HE)), other reported partial agonist-antagonists (*i.e.*, RU39411 (RU): Gottardis *et al.*, Cancer Res. 49:4090-4093 (1989); 119010 (119): Nishino *et al.*, J. Endocrinol. 130:409-414 (1991); centchroman (Cen): Hall, BBRC 216:662-668 (1995)), and a pure antagonist (*i.e.*, ICI164384: Wakeling *et al.*, J. Endocrinol. 112:R7-R10 (1987)) for these initial studies in order to determine whether compounds with different *in vivo* actions would be distinguishable using any of these assays.

Since 1.0 μ M concentrations of compound were shown to elicit a maximal response in most of the assays, all compounds were tested at 1.0 μ M. In some cases, 10 μ M concentrations were also tested. The ability of a compound to alter steady state levels of mRNA corresponding to each gene was quantitated by Northern, slot blot, or RT-PCR analysis as described herein (Table 1). The average fold-increase in mRNA levels elicited by either E2 or EE for each gene/cell assay is provided in the third column. Compounds that elicited a response in a particular assay are designated with a (+); those that showed no effect are designated with a (-). Analysis of 27 different gene/cell combinations with these nine compounds (to generate a GEF for each compound) revealed that most of the assays provided redundant information (seen as the same pattern of activity across the series of compounds in Table 1); but, five distinct activity patterns across this set of compounds were discernable among all these gene/cell combinations, as indicated by the roman numerals I-V on the rights side of Table 1. It is of interest that pattern I is found in most of the cell types tested, but the other patterns

(particularly pattern II) may show a cell-type preference. Such data emphasizes the value of using different cells as well as different genes in carrying out these analyses. Also evident in Table 1 is the fact that compounds can have differential abilities to activate the same gene (*e.g.*, 52kD) depending upon the cell (*e.g.*, ZR75-1 compared to BG-1).

Table 1. ACTIVITY of SELECTED ER LIGANDS in MODULATING GENE EXPRESSION

| Gene | Cell Line | Fold Effect+E | E2 | EE | 2HE | HT | Ral | FU | 119 | Cen | ICI | |
|----------|-----------|---------------|----|----|-----|----|-----|----|-----|-----|-----|-----|
| FR | GH3 | 12 | nd | + | + | - | - | - | - | - | - | I |
| FR | MCF7 | 10 | + | nd | + | - | - | - | nd | nd | - | |
| FR | B5-ER | >20 | + | + | + | - | - | - | - | - | - | |
| FR | ZR75-1 | 15 | + | + | + | - | - | - | - | - | - | |
| FR | Ishikawa | 15 | + | nd | + | - | - | nd | nd | - | - | |
| FR | BG-1 | 25 | + | + | + | - | - | - | - | - | - | |
| GH | GH3 | 5 | nd | + | + | - | - | - | nd | nd | - | |
| 52kD | MCF7 | 5 | + | nd | + | - | - | - | nd | nd | - | |
| 52kD | B5-ER | 9 | + | + | + | - | - | - | - | - | - | |
| IGFBP-1 | HepG2-ER1 | 5 | + | + | + | - | - | - | - | - | - | |
| pS2 | MDA361 | 10 | + | + | + | - | - | - | - | nd | - | |
| pS2 | MCF7 | 10 | + | nd | + | - | - | - | nd | nd | - | |
| pS2 | BG-1 | >20 | + | + | + | - | - | - | - | - | - | |
| Amphireg | MDA361 | 8 | + | + | + | - | - | - | - | nd | - | |
| 52kD | BG-1 | 5 | + | + | + | - | - | - | - | - | - | |
| VEGF | BG-1 | 3 | + | + | + | - | - | - | - | - | - | |
| pS2 | ZR75-1 | 5 | + | + | + | - | - | - | - | - | - | |
| TRHR | GH3 | 5 | nd | + | + | + | + | + | + | + | - | II |
| PRL | GH3 | 39 | + | + | + | + | + | + | + | + | - | |
| pS2 | MDA-ER | ~450 | + | + | + | + | + | + | + | + | - | |
| TGFalpha | B5-ER | 10 | + | + | + | + | - | + | - | + | - | |
| TGFalpha | MDA-ER | 11 | + | + | + | + | - | + | - | + | - | |
| 52kD | ZR75-1 | 6 | + | + | + | + | - | + | - | + | - | III |
| OBG | HepG2-ER2 | 4 | + | + | + | + | - | + | - | + | - | |
| pS2 | B5-ER | >20 | + | + | + | + | - | + | - | + | - | |
| FR | MDA-ER | ~500 | + | + | + | + | - | + | + | + | - | IV |
| IGFBP-1 | Fe33 | 12 | + | + | - | + | + | + | - | - | - | V |

Summary of the maximal responses of each gene/cell combination (*i.e.* assay) to compound treatment. +, active compounds; -, inactive compounds; nd, not determined. The cell lines listed were treated with the indicated compounds, total RNA was isolated, and analysed for modulation of expression of the listed genes as described in M&M. The maximal average gene expression response of each cell line following E2 or EE treatment is provided in the third column (*i.e.* Fold Effect + E). Each assay can be grouped according to their response to compound treatment into the classes shown at the right side (*i.e.* I-V).

Furthermore, the same compounds can have different activities on different genes within the same cell (*e.g.*, PR compared to pS2 or TGF- α in the MDA-ER cells).

Thus, for this selected compound set, five non-redundant "informative" assays, *i.e.*, those whose combined use enable the discrimination of compounds into different classes were identified in the twenty-seven assays analyzed. It is noteworthy that not all five assay types (patterns) were equally represented. The predominant assay type showed responsiveness only to estradiol derivatives (*i.e.* EE and 2HE) whereas the least frequently identified patterns (corresponding to the assays that score Ral and 119) were observed only 4 times. Thus, of the estrogen response assays used herein, a subset thereof chosen randomly would comprise at least 15 and preferably as many as 20 assays for use in the GEF screen. The statistical probability of identifying raloxifene as an active compound in such a screen would be 96% if 20 assays are employed, 91% if 15 assays are used, and 80% if only 10 are analyzed (Snedecor *et al.*, Statistical Methods, 8th ed. Iowa State University Press, Ames, Iowa, Chapter 7, (1989)).

To simplify the GEF screen, additional studies were performed to determine which of the redundant assays was most amenable to screening strategies (*e.g.*, highest reproducibility and extent of change relative to control). The IGFBP-1/Fe33 gene-cell combination (representing pattern V) was not employed in further studies (due to difficulties interpreting data in these liver carcinoma-derived cells, where drug-metabolizing activity is significant). The chosen representative assays for subsequent studies are shown in Table 2. This representation of the data shows that each compound is identified by a specific GEF based upon the activity elicited in each of the four assays (seen as + and - pattern of activity in the column underneath each compound). In this manner, compounds with identical GEFs were grouped together and were distinguishable from those with different GEFs. For example, E2, EE, and 2HE were placed in one group (#1 in Table 2) and HT and RU in another (#2). Of utmost importance was the observed difference between E2, Ral, and HT, which indicated that these assays are successful in discriminating among compounds with distinct *in vivo* pharmacologies.

Table 2. ER LIGAND CLASSIFICATION by GEF

| Gene | Cell Line | E2 | EE | 2HE | HT | RU | Ral | 119 | Cen | ICI |
|----------|-----------|----|----|-----|----|----|-----|-----|-----|-----|
| PR | BG-1 | + | + | + | - | - | - | - | - | - |
| PRL | GH3 | + | + | + | + | + | + | + | + | - |
| TGFalpha | MDA-ER | + | + | + | + | + | - | - | + | - |
| PR | MDA-ER | + | + | + | + | + | - | + | + | - |
| Group # | | 1 | 1 | 1 | 2 | 2 | 3 | 4 | 2 | 5 |

3. Classification of additional compounds using selected gene/cell assays

This method of classification was employed to separate an additional thirty compounds, many of which are structurally related to the first nine compounds tested. Compounds E1 (estrone), E3 (estriol), DHE (17 α -dihydroequilen), DHEN (17 α -dihydroequilenin), ZK182491 and ZK155843 are derivatives of either 17 α -estradiol (17 α -E2) or 17 β -estradiol. Compounds ZK166780, ZK166781, ZK167466, ZK167957, and ZK180686 are 11 β -substituted 17 β -estradiol derivatives related to RU39411. Compounds HT, ZK186275, ZK183819, ZK182956, and ZK183955 are tamoxifen derivatives. Compounds ZK185157 and ICI182780 are related to the pure steroidal antagonist, ICI164384. Compounds ZK182254, ZK186217, and raloxifene are benzothiophenes. Compounds ZK183659, ZK22496, and ZK185704 are structurally related (*i.e.*, contain a cyclophenyl moiety). Compound ZK167502 is a naphthalene derivative and coumestrol is a phytoestrogen (Price *et al.*, Food Addit. Contam. 2:73-106 (1985)). Many of these had been previously classified as agonists, partial agonists, or antagonists of the ER through assays of ER binding and transcriptional activation. In these experiments, compounds were scored using three activity levels (*i.e.*, inactive, partially active as <50% of the E2 response, fully active as >50% of the E2 response). As is evident from Table 3, the compounds could be divided into ten groups by this analysis (see Table 3). This separation of compounds is not based primarily upon chemical structure as indicated by the results with the compounds that are related to RU39411 (*i.e.*, ZK166780, ZK166781, ZK167466, ZK167957, and ZK180686). These six compounds are split into 3 different classes based on their GEFs.

Table 3. RESULTS OF GEF ANALYSIS

| COMPOUND | PR/BG-1 | PRL/GH3 | TGF α /MDA-ER | PR/MDA-ER | GROUP |
|-----------------|---------|---------|----------------------|-----------|-------|
| E2 | ++ | ++ | ++ | ++ | 1 |
| EE | ++ | ++ | ++ | ++ | |
| E1 | ++ | ++ | ++ | ++ | |
| E3 | ++ | ++ | ++ | ++ | |
| Coumestrol | ++ | ++ | ++ | ++ | |
| 167502 | ++ | ++ | ++ | ++ | |
| 2HE | ++ | ++ | ++ | ++ | |
| DHE | + | ++ | ++ | ++ | 2 |
| DHEN | + | ++ | ++ | ++ | |
| 182491 | + | ++ | ++ | ++ | |
| 155843 | ++ | + | ++ | ++ | 3 |
| 17 α -E2 | ++ | + | ++ | ++ | |
| 22496 | ++ | + | ++ | ++ | |
| 166780 | + | + | ++ | + | 4 |
| 166781 | + | + | ++ | + | |
| RU39411 | - | + | ++ | + | 5 |
| HT | - | + | ++ | + | |
| Centchroman | - | + | + | + | 6 |
| Tamox | - | + | + | + | |
| 186275 | - | + | + | + | |
| 182254 | - | + | + | + | |
| 185704 | - | + | + | + | |
| 183955 | - | + | - | + | 7 |
| 119010 | - | + | - | + | |
| Ralox | - | + | - | - | 8 |
| 186217 | - | + | - | - | |
| 183819 | - | + | - | - | |
| 167466 | - | - | - | + | 9 |
| 167957 | - | - | - | + | |
| 185157 | - | - | - | + | |
| 180686 | - | - | - | + | |
| 182956 | - | - | - | - | 10 |
| 183659 | - | - | - | - | |
| ICI164384 | - | - | - | - | |
| ICI182780 | - | - | - | - | |
| progesterone | - | - | - | - | |
| RU486 | - | - | - | - | |
| resveratrol | - | - | - | - | |
| dexamethasone | - | - | - | - | |
| ph n l red | - | - | - | - | |

Data represent the average maximal response (at concentrations up to 10 μ M) of at least three individual experiments with duplicate determinations. Activity ++, >50% E2; +, <50%; -, inactive.

4. Determination of the predictive ability of the GEF classification for *in vivo* effects

Included in the compounds tested in the section above were standards (*i.e.*, E2, tamoxifen, raloxifene, ICI 164384) with reported distinguishable *in vivo* profiles. E2, tamoxifen, and raloxifene, but not ICI, have "estrogenic" effects on the bone and cardiovascular system in experimental and/or clinical studies (*i.e.*, they are effective in attenuating atherosclerotic lesion formation tamoxifen: Williams *et al.*, Arterioscler. Thromb. Vasc. Biol. 17:403-408 (1997); raloxifene: Bjarnason *et al.*, Circulation 96:1964-1969 (1997) and/or protecting against ovariectomy-induced bone loss (tamoxifen: Love *et al.*, N. Engl. J. Med. 326:852-856 (1992); raloxifene: Black *et al.*, J. Clin. Invest. 93:63-69 (1994)). Yet, E2 and tamoxifen were readily distinguishable from raloxifene in their greater potency in eliciting uterotrophic effects (Sato *et al.*, FASEB J. 10:905-912 (1996) and Table 4), thereby implying that raloxifene has tissue-selective actions *in vivo*. Through our analysis of gene expression patterns, we found that these four compounds have different GEFs that place them in separate groups (Tables 2 and 3). These data support the idea that compounds with selective *in vivo* effects be distinguished by different gene expression profiles (GEFs) *in vitro*.

Of particular interest was the group of compounds including ZK167466 (Group 9, Table 3). Like the raloxifene group (Group 8), these compounds exhibited activity in only one GEF assay. To determine whether the co-classification of these compounds predicted similar *in vivo* pharmacology, they were tested *in vivo* for uterotrophic activity as well as their ability to reduce the loss in bone mass caused by decreased circulating levels of estrogens (*i.e.*, induced experimentally by ovariectomy). Table 4 compares the activity of this group of compounds to E2, Tam, Ral, and ICI. All four of the group 9 compounds were different from the others in both assays. They showed either no or only weakly stimulatory effects (depicted as - or -/+ in Table 4) in promoting endometrial thickening (*i.e.*, uterotrophic effect). Three of them are significantly effective in the "bone protection" assay that predicts efficacy against osteoporosis (Table 4). These data indicate that this GEF profile predicts a novel selective compound class (*i.e.*, one with bone-protective effects and little or no uterotrophic response), which could not have been identified (separated from the other "partial agonists") with the existing *in vitro* screening methods.

Table 4. IN VITRO ANALYSIS: GEF IN VIVO ANALYSIS

| COMPOUND | PR/BG-1 | PRL/GH3 | TGF α MDA-ER | PRMDA-ER | Uterotropic Response | Bone Protection |
|-----------|---------|---------|---------------------|----------|----------------------|-----------------|
| E2 | ++ | ++ | ++ | ++ | ++ | ++ |
| Tamox | - | + | + | + | ++ | ++ |
| Ralox | - | + | - | - | + | ++ |
| 167466 | - | - | - | + | -/+ | + |
| 167957 | - | - | - | + | - | + |
| 185157 | - | - | - | + | -/+ | -/+ |
| 180686 | - | - | - | + | -/+ | + |
| ICI164384 | - | - | - | - | - | ND |

Data for the In Vitro Analysis (from Table 3) is compared with the activity of the same compounds in In Vivo Studies. The average maximal response relative to control is shown for each compound. Uterotropic response is the change in endometrial epithelial cell height elicited by compound treatment. Bone protection is determined from the % loss in bone mineral density (quantitated by pQCT) of control and compound-treated animals. ++, >70% E2; + 35-70% E2; -/+ 10-35% E2; -, < 10% E2; ND, not determined.

C. Materials and Methods

1. Cell Culture and Compound Treatment

MDA-231 ER transfectant E-28 cells were routinely cultured in phenol red-free alpha-modified minimal essential medium (MEM Gibco BRL; Gaithersburg, MD) supplemented with 1 millimolar (mM) HEPES, 2mM glutamine, 0.1 mM MEM non-essential amino acids, 1.0 mM sodium pyruvate, 50 µg/ml gentamicin (all from Gibco), 1.0 microgram/milliliter (µg/ml) insulin (Sigma; St. Louis, MO), and 5% DCC-treated FBS (Intergen). Cells were plated at approximately 40% confluency (1.5 x 10⁶/plate) in 150 mm culture dishes. Following an overnight cell attachment, the medium was changed to include 0.2% ethanol or the test compounds and cultured for an additional 48 hours (h).

GH3 rat pituitary cells were routinely cultured in DMEM-F10 (1:1) medium containing 12.5% horse serum, 2.5% FBS, 25 mM Hepes, 2 mM L-glutamine, and 50 µg/ml gentamicin sulfate at 37°C, 5% CO₂. Under these conditions, the cells were partially adherent, and both adherent and non-adherent cells were maintained during the passaging of the cells. For the measurement of mRNA expression, cells were seeded (10⁶/100 mm dish) in culture medium without phenol red and containing DCC-treated serum. After 3 days, the medium was changed to one containing 0.2% ethanol or the test compounds, and the cells were further incubated for 2 days.

BG-1 human ovarian carcinoma cells (Geisinger *et al.*, Cancer 63:280-288 (1989)) were cultured in DMEM:F12(1:1) medium containing 10% FBS, 2 mM L-glutamine and 50 µg/ml gentamicin sulfate. For the measurement of mRNA expression levels, cells were cultured for 24h in phenol red-free medium containing 5% DCC-treated FBS prior to plating in the same medium at a density of 2 x 10⁶/150 mm plate. The following day, the medium was changed to include 0.2% ethanol or the test compounds and cultured for an additional 72h.

ZR75-1, MCF7, and MDA361 human breast carcinoma cell lines were routinely cultured in alpha-modified MEM supplemented with 1 mM HEPES, 2 mM glutamine, 0.1 mM MEM non-essential amino acids, 1.0 mM sodium pyruvate, 50 µg/ml gentamicin, 1.0 µg/ml insulin, and 10% FBS. Cells were plated (ZR75-1: 1.5 x 10⁶/p100; MCF7: 2 x 10⁶/p150; MDA361: 5 x 10⁶/p100) in phenol red and insulin-free media containing 5% FBS-DCC for the assays. Following an overnight cell attachment, the medium was changed to include 0.2% ethanol or the test compounds and cultured for an additional 24h (ZR75-1), 48h (MDA361), or 72h (MCF7).

The HepG2 human hepatocarcinoma cells, stably transfected with ER (clones ER1 and ER2), were cultured in EMEM (GIBCO), supplemented with 1 mM HEPES, 2 mM glutamine, 0.1 mM MEM non-essential amino acids, 1.0 mM sodium pyruvate, 50 μ g/ml gentamicin, and 10% FBS. Ishikawa human endometrial carcinoma cells were cultured in EMEM with 2 mM glutamine, 50 μ g/ml gentamicin, and 10% FBS. Fe33 (ER-transfected FTO-2B rat hepatoma cells) were maintained in DMEM-Ham's F12 (1:1) without phenol red containing 10% DCC-FBS on 0.1% gelatin coated Petri dishes. All cells were plated (HepG2-ER: 4×10^6 /p100; Ishikawa: 2×10^6 /p150; Fe33: 2.5×10^5 /p150) in phenol red and insulin-free media containing 5% FBS-DCC for the assays. Following an overnight cell attachment, the medium was changed to include 0.2% ethanol or the test compounds and cultured for an additional 72h.

The ER-transfected human mammary epithelial cells (B5-ER) were maintained and assayed for gene expression changes according to protocols previously described (Zajchowski *et al.*, Mol. Endocrinol. 5:1613-1623 (1991)). Compound or vehicle treatment was for 72h.

17β -estradiol, 17α -ethinyl estradiol, estrone, estriol, progesterone, dexamethasone, phenol red were purchased from Sigma Biochemicals (St. Louis, MO). All other compounds were synthesized at Schering AG (Berlin). Stock solutions (10 mM) of all the chemicals were prepared in DMSO and diluted in ethanol for the assays.

2. RNA Isolation and Slot Blot Analyses

At the end of the compound treatment time, cell monolayers were harvested into Ultraspec (Biotecx Laboratories, Houston, TX) or RNeasy (Qiagen Inc., Santa Clara, CA) RNA isolation reagent and processed according to the manufacturer's suggested protocol. Total RNA (MDA-231 ER: 10 μ g; GH3: 1.0 μ g) was spotted onto a Zetaprobe-GT nylon membrane using a 48-well slot blot apparatus attached to a vacuum manifold. Total RNA (20 μ g) from treated and untreated samples of all of the other cell lines was evaluated by Northern blot analysis. Hybridization of the membranes to 32 P-dCTP labeled probes was carried out as previously described. Quantitation of the specific hybridization in each spot by subtracting non-specific background detected in a negative control for each mRNA was performed using a Fuji phosphorimager; the ratio of the signal intensities in compound-treated samples relative to controls provided the value for fold-change used in the assessment of the compound activity for each particular assay. Changes in mRNA levels greater than or equal to 2-fold were scored as positive.

3. Progesterone Receptor Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR)

All RNA samples were diluted to 20 ng/ μ l in DEPC-treated water. RT PCR was performed using 100 ng total RNA. The reaction mixtures contained 5 units
5 rTth DNA Polymerase (Perkin Elmer; Foster City, CA), 1X EZ buffer (Perkin Elmer; Foster City, CA), 2.5 mM Mn(OAc)₂, 300 μ M dNTP's (mix from Pharmacia; Alameda, CA) and 10 pmol of each biotinylated primer in a final volume of 50 μ l. PCR primers PR#1 (5' GTC AGT GGR CAG ATG CTR TAT TT), PR#2 (5'-11C TTC AGA CAT CAT TTC YGG AAA TTC) were synthesized by Synthetic Genetics (San Diego, CA).
10 Amplification consisted of a 30 minute RT step at 60°C immediately followed by 33 cycles of a two step PCR reaction (95°C for 15 seconds, 60°C for 45 seconds) and a final 7 minute extension at 60 C in a Perkin Elmer 9600. Following PCR, 1/20 reaction volume is removed and quantitated using streptavidin-coated 96-well microplates and oligonucleotide probes specific for the PCR target. The probe is coupled to either HRP
15 or AP and addition of either colorimetric (HRP) or chemiluminescent (AP) substrates permits quantitation of 300-500 initial copies of specific RNA template in a 20-100 ng total RNA sample. *In vitro*-transcribed PR mRNA was used to generate standard curves (calculated by non-linear regression analysis using a four parameter sigmoidal plot) for quantitation of the amount of PR mRNA in each reaction. Changes in mRNA levels
20 were scored as positive if they were greater than or equal to 3-fold.

4. Uterine Histomorphometric Analysis

For determination of uterotrophic activity, immature, 19-21 day old female Sprague-Dawley rats, weighing 35-50 g. were given daily subcutaneous injections for
25 three days with compounds or vehicle alone. The compounds were dissolved in a vehicle consisting of 10% ethanol in arachis oil or a mixture of benzylbenzoate/castor oil (1:4). On day 4, the animals were weighed and euthanized by carbon dioxide asphyxiation. The uteri were excised and placed in neutral buffered 3.7% formaldehyde for a minimum of 24 hours. The uteri were then embedded in paraffin, cut into 4- μ m transverse
30 sections, and stained with hematoxylin and eosin and the sections evaluated for luminal epithelium cell height as described by Branham *et al.* (Branham *et al.*, Biol. Reprod. 53:863-872 (1995)). The difference in epithelial cell height between the estrogen (0.3 μ g 17 β -estradiol/animal) and vehicle-treated groups was calculated and expressed as 100%.

The activity of the compound of interest as a percent of 17 β -estradiol was calculated according to the following formula:

$$x = \frac{100\%[\text{height}(\text{test compound}) - \text{height}(\text{vehicle})]}{\text{height}(17\beta\text{-estradiol}) - \text{height}(\text{vehicle})}$$

5. Bone Mineral Density Measurement

For determination of efficacy in preventing bone loss, 3 month old female rats (Sprague Dawley) were ovariectomized (ovx) and treated immediately after surgery. Compounds were applied once daily s.c. in benzyl benzoate/castor oil (1:4) or arachis oil/ethanol (95:5). Control groups (sham/ovx - treated with vehicle) and treatment groups consisted of 6 animals each. 4 weeks after surgery animals were sacrificed and the left and right tibia were processed for bone mineral density measurements. Bone mineral density (BMD) was measured in the secondary spongiosa of the proximal tibia by pQCT (peripheral quantitative computed tomography). Results are expressed in percent protection from bone loss. Bone protection was expressed relative to the effects of estrogen (0.3 μ g 17 β -estradiol/kg) according to the following formula:

$$x = \frac{100\%[\text{BMD}(\text{test compound}) - \text{BMD}(\text{vehicle})]}{\text{BMD}(17\beta\text{-estradiol}) - \text{BMD}(\text{vehicle})}$$

II. Screening for an Interferon- β (IFN β) Mimetic

A. Background

IFN β has efficacy in the treatment of Multiple Sclerosis (MS) (The IFN β Multiple Sclerosis Study Group Neurology 43:655-661 (1993)). The precise mechanism by which IFN β elicits its therapeutic efficacy is unknown. However, a great deal of knowledge exists concerning the signal transduction pathways modulated by IFN β ; as a ligand, IFN β directly interacts with its receptor to induce phosphorylation of a number of signal transducing proteins (STATs (Ihle, Nature 377:591-594 (1995)) and eventually direct specific changes in gene expression (Darnell *et al.*, Science 264:1415-1421 (1994)). A homologous member of the same family of cytokines, IFN α , is capable of binding the same receptor protein yet cannot be used in the treatment of MS due to its unacceptable side effect profile. Another interferon, IFN γ , shares some of IFN β 's effects on gene expression, yet actually exacerbates the symptoms of MS (Panitch *et al.*, J. Neuroimmunol. 46:155-164 (1993)). Therefore, differences in the biological effects of

these three ligands can be exploited in developing screens to identify selective IFN β mimetics that might be more efficacious and have better tolerability than IFN β itself.

Animal models to test drug efficacy in ameliorating the severity of this disease exist (*i.e.*, Experimental Autoimmune Encephalitis (EAE) or T cell transfer EAE model).

B. Cell selection and gene identification

Cells employed in these studies can be representative of known or suspected IFN β -responsive tissues (*e.g.*, B cells (*e.g.*, Daudi), T cells (*e.g.*, Jurkat), glioblastoma (*e.g.*, T98G), carcinoma (A549), and astrocytes (*e.g.*, CH235)). RNA is prepared from candidate cell lines that have been treated with IFN β and used to estimate the number of differentially expressed sequences by hybridizing probes prepared from this RNA on microarrays containing 100 or more pre-selected cDNAs, such as the Atlas cDNA Arrays (*i.e.*, Clontech). The cell lines that show the largest number of differentially expressed sequences are chosen for studies to identify IFN β -responsive genes. Technically, this can be approached through any available differential gene expression screening strategy (*e.g.*, DD-PCR, subtractive hybridization libraries, *etc.*). Subsequent to identification of the differentially-expressed genes, limited optimization is preferred to determine whether conditions such as time of treatment can enhance the extent of mRNA change relative to control. Conditions amenable to analysis of the largest number of genes are used.

C. Assay characterization

For each cell line, genes that show significant regulation (preferably at least a 5-fold increase or decrease from basal level) are used in screens with a set of compounds known to have different, but overlapping effects in common with IFN β (*e.g.*, IFN α , IFN γ , IL-8, IL-12). This evaluation can be carried out by arraying the cDNAs for these candidate genes and using RNA isolated from each of the compound-treated cells to prepare hybridization probes. Responsive genes are evaluated for the response to each compound. An exemplary set of one or more genes, including gene/cell combinations, responds only to IFN β , another group of genes responds to both IFN α and β , another with IL-8, IFN γ , and IFN β , *etc.*

D. Assay selection

The "best" gene/cell combination (greatest fold response and signal-to-noise ratio for detection; gene expression measurable in cell line where other "informative" genes are measured) from each group of genes is chosen for the compound screen. Internal control genes are designated in the cell line to be used as indicators of cytotoxicity (*e.g.*, *gadd45*, *hsp 70*).

E. Screening

A test compound library is screened for those test compounds which are specific modulators of IFN-responsive genes using a scoring method of active and inactive. The "active" hits are those that elicit changes in gene expression significantly above the background variance of the specific assay. Test compounds are then grouped according to their GEF and re-tested to determine the EC_{50} for representative compounds.

At this stage in the generation of a GEF that will be predictive for *in vivo* efficacy, it may not be clear how close to the GEF of IFN β a "hit" will need to be in order to have IFN-like activity *in vivo*. To estimate this, test compounds that showed activity in the greatest number of assays (*i.e.* gene/cell combinations) are tested in a cell-based assay for IFN responses (*e.g.*, anti-viral effects) prior to *in vivo* testing. This screen is employed as a way of sorting through GEFs to determine whether "hits" with activity in very few IFN-response assays have IFN-like activity. If none of the hits that are active in multiple GEF assays show activity in the bioassay, compounds are preferably screened in combination with each other to determine their GEF upon co-treatment. Combinations of compounds that generate new GEFs closer to that of IFN β are subsequently tested for *in vitro* activity in the bioassay.

Representative compounds are selected for *in vivo* evaluation based upon their activity in *in vitro* bioassays, potency in the GEF assays, and other available information. If any "hits" meet criteria for *in vivo* testing, they are evaluated for efficacy in the EAE model. If not, additional compound sources can be screened, or weak "hits" can be optimized against their GEF to find more potent compounds before testing in animal models.

F. Selectivity testing and "selective" GEF determination

The GEF profile determined in the previous step can be used directly as a means of optimizing "lead" or representative best candidate compounds. At this stage of analysis, EC50s and maximal responses for the derivative compounds for each assay are considered.

The "lead" compound(s) is usually tested for adverse, undesirable effects in appropriate biological models (*e.g.*, induction of fever, testable in a rabbit pyrogenicity assay). If there are "lead" compounds that have different GEFs, the GEF corresponding to the "lead" which has little or no activity in this assay is used for further optimization. If, however, none of the "lead" compounds meet the selectivity requirements for the desired drug, it may be necessary to incorporate additional assays into the screening panel and re-test all of the bioactive "hits"; in this new screen, compounds within the previously designated GEF classes may be differentiated from each other by these new assays (*i.e.*, due to a different GEF that is now discovered). In that case, additional *in vivo* evaluation is necessary to validate the predictability of the new GEF for *in vivo* efficacy and selectivity.

III. Identification of a p53 Mimetic for Cancer Treatment

A. Background

Mutation or deletions of the p53 tumor suppressor gene are prevalent in many human cancers (Hollstein *et al.*, Science 253:49 (1991); Weinberg, Science 254:1138 (1991)). Studies during the last decade have elucidated the dominant role that this protein plays in maintaining the normal balance between cell proliferation and death. Most importantly, experimental evidence from both *in vitro* and *in vivo* studies has demonstrated the feasibility of p53 protein replacement as a treatment for cancer (Wills *et al.*, Hum. Gen. Ther. 5:1079-1088 (1994)).

In addition to its transcriptional regulatory activities, p53 has been shown to influence DNA replication and repair as well as apoptotic signaling pathways. A profile of the changes in gene expression that result from the expression of wild type (WT) p53 in a cancer cell will be used in the application presented here as a tool to search for compounds that mimic the activities of p53. The existence of expression systems that enable investigator-control of protein expression (*e.g.*, lac or tet-inducible systems) as well as temperature sensitive (ts) p53 proteins and a number of p53 mutants enhance the suitability of this system for drug-screening efforts.

B. Cells and Genes

Cancer cell lines which have been stably modified (*e.g.*, by transfection or transduction techniques) to enable regulatable expression of the p53 WT or mutant variants are used to identify p53-dependent genes. These studies would preferably be performed in a p53 null cell background, although this criterion is not absolute. Any of the methods described in previous examples can be employed to identify candidate p53-responsive genes. RNA for this analysis is isolated from cells cultured under conditions where (1) the expression of the p53 protein is on or off (*e.g.*, in an inducible expression system) or (2) the active vs. inactive form of the p53 protein is present (*e.g.*, for a temperature sensitive p53 protein or for WT vs. mutant proteins).

In this example, (1) the effector compound is a 53kD protein (*i.e.*, p53) and not a small molecule (*i.e.*, estradiol) or a polypeptide ligand (*i.e.*, IFN- β) and (2) the search is for an alternative effector molecule(s) which elicits the same *in vivo* effects as p53, not a more selective or efficacious molecule. In this regard, it is important to note that a successful p53 mimetic could be a combination of compounds, each of which perform a "subset" of the essential p53 functions. In the previous instances, the cell line(s) which showed the greatest number of changes in response to the reference compound was chosen for the identification of responsive genes. In this case, a minimal set of gene/cell readouts that are predictive of p53's tumor suppressive function is the desired outcome of the assay selection step. Therefore, the initial gene identification approach will evaluate several different tumor cell lines whose tumorigenicity is suppressed by p53 introduction/activation. The p53-responsive assays that are shared by all of these cells are selected for further evaluation.

C. Assay characterization and selection

An additional, but not essential, method for choosing the appropriate assays is to evaluate the expression of candidate genes following induction of the WT p53 compared to its mutated versions. Genes which are regulated by truncated or mutated p53 proteins that retain their tumor suppressor function are useful in a p53 mimetic screen since they are markers of desirable p53 functions; genes which continue to be regulated by mutant versions of p53 that are inactive in tumor suppression would be eliminated from the screen or used as "non-selective" assays. The choice of assays to be used as read-outs of "cytotoxicity" may differ in this screen from those applications

described above, since some of the targets of p53 may be genes like gadd45; the assays which do not respond to p53 can be retained as "cytotoxicity" readouts.

Evaluation of gene expression patterns elicited by compounds will be similar to other searches. "Hits" will be grouped according to their GEF and re-tested to determine EC₅₀ for each active assay.

D. Preliminary cell-based assays

The "hits" can initially be tested in *in vitro* assays for proliferation (*e.g.*, measured by ³H-thymidine uptake), anchorage-independent growth (*e.g.*, soft agar assays), and apoptosis (*e.g.*, measured by DNA-laddering induced upon exposure to radiation in the presence of the compound). This preliminary evaluation will further define the GEF that predicts activity in tumor suppression (as measured by the *in vitro* surrogate assays). The *in vitro* systems can be also used to evaluate efficacy of combinations of "hits" that may synergize to generate a GEF that predicts tumor suppressor function.

E. In vivo evaluation

Representative compounds are selected for *in vivo* evaluation based upon their activity in *in vitro* bioassays, potency in the GEF assays, and other available information. The efficacy of compounds in suppressing the growth of human tumor xenografts in nude, athymic mice will be assessed as a measure of tumor-suppressive activity. Positive controls for this study are the same tumor cells which are engineered to express an inducible p53 protein, which enables regulation of tumor growth *in vivo*.

F. GEF definition and lead compound optimization

The GEF profile that correlates with *in vitro* and *in vivo* efficacy can be used directly as a means of optimizing "lead" compounds. This is a preferred step for any combinations of compounds that are active in the *in vitro* bioassays, since the combination therapy may be difficult to evaluate in *in vivo* assays due to possible pharmacokinetic differences of the components of the mixture. At this stage of analysis, EC₅₀s and maximal responses for the derivative compounds for each assay are considered.

Depending upon the selectivity requirements for the desired drug, it may be useful to incorporate additional assays into the screening panel at this stage. In that

case, additional *in vivo* evaluation is necessary to validate the predictability of the new GEF for *in vivo* efficacy and selectivity.

IV. Identification of Agents that Block Cell Invasion for Cancer Therapy

Therapeutic agents that prevent the progression of primary cancer to the metastatic stage are important members of the arsenal of anti-cancer drugs. Different aspects of the process by which a cancer cell enters the bloodstream, leaves it, and re-establishes itself at a distant site are potential targets for anti-metastatic drugs. However, there is a paucity of *in vitro* and *in vivo* models that predict the metastasis-forming ability of human cancer cells; this makes the identification of anti-metastatic agents particularly challenging.

A critical aspect in this progression is the process by which cells pass through the endothelial lining of the blood vessel and invade into the surrounding stroma. Cell invasion through a reconstituted basement membrane (*e.g.* Matrigel) can be employed as an *in vitro* surrogate for the *in vivo* event. The assay, however, is not readily adaptable to the screening of large compound libraries. The GEF methodology can be used to develop a screen for agents that block or decrease cell invasion and/or metastasis.

Rather than employing a reference compound for identification of gene expression differences, the genes for this screen are identified by comparing reference states. Exemplary reference states may include, but are not limited to the following: invasive vs. non-invasive cell lines, normal vs. invasive carcinoma tissue, or two histopathologically-staged malignant tissues (*e.g.*, prostatic carcinomas of Gleason Grades III and IV).

A. Cells and Genes

Both cells and tissue specimens which represent various stages in cancer progression (*e.g.* from normal to highly invasive or metastatic) are used as sources of RNA. An exemplary set of cell lines or strains for studies of breast cancer progression is based, for example, on reported *in vitro* invasive properties (*e.g.*, normal human mammary epithelial cells, immortal MCF10A or 184B5, poorly invasive MCF7, ZR75-1, MDA468, moderately invasive MDA435, and highly invasive MDA231 or BT549 (available from ATCC, Rockville, MD). Tissue samples can include human xenografts from immunodeficient animals, biopsies that have been dissected by a pathologist to

specifically include tumor, normal, and invasive material or similarly characterized cells generated, for example, by Laser Capture Microdissection (Emmert-Buck *et al.*, Science 274: 998-1001 (1996)). Although there is scientific rationale for the comparison to be made amongst cells and biopsy specimens derived from the same tissue of origin, this is not required because a process common to the metastasis of different cancer types could be targeted by deriving a screen using cells and biopsies from other tissues.

Several approaches can be taken to determine the gene expression differences and similarities among these RNA samples. The RNA isolated from the normal and the most invasive cells (or biopsies) can be compared using methods described above for identifying differences between treated and untreated cells (*e.g.* DD-PCR, subtractive cDNA libraries, high density cDNA arrays). Pooled samples from normal vs. tumor cell lines or specimens representing different stages of cancer progression may also be used to generate this gene expression comparison and are, in fact, preferred because of the greater pool of differentially expressed sequences that is likely to be generated. This is particularly important with regard to the tumor cells, since it is known that there is individual variability in tumors; these differences are likely to be reflected in different gene expression profiles.

The genes that are differentially expressed between normal and highly invasive cells are selected for further evaluation.

B. Assay characterization and selection

Genes identified as differentially expressed in the first step are assessed for inclusion in the GEF based upon their expression in the cells being considered for use in the screening process. For example, if the initial gene identification was carried out using RNA isolated from tissue specimens and not cell culture material, some genes expressed *in vivo* may not be similarly expressed or regulated in the culture environment. Preferably cell lines which express the greatest number and the highest levels of mRNA for the differentially expressed genes would be chosen for the GEF assays.

In the process of evaluating the expression of the candidate genes in normal vs. invasive cultured cells, it is also desirable to test their relative expression in tumor cells that are either not invasive or poorly invasive. By comparing the gene expression patterns in these cells, a subset of the genes can be identified that is commonly modulated in only invasive cells or in the majority of the invasive cell lines tested. This subset will be especially informative for inclusion in the GEF.

In some embodiments, regulation of expression of any of the candidate genes by agents that are reported to modulate cancer cell invasion (*e.g.* TGF β , metastasis suppressor nm23, anti-Ha-ras ribozymes) is determined. The genes whose expression is affected by these agents are then included in the GEF.

5 The "best" assays (*e.g.* gene/ cell combination with greatest fold response and signal-to-noise ratio for detection) are chosen for the compound screen. Appropriate genes to be used as indicators of cytotoxicity (*e.g.* gadd45, hsp 70) or as internal controls (*e.g.*, GAPDH) are also incorporated into the GEF.

10 C. Compound screening

Evaluation of gene expression patterns elicited by compounds is similar to other searches described above. "Hits" are grouped according to their GEF and re-tested to determine EC₅₀ for activity in each assay.

15 D. Preliminary cell-based assays

The "hits" are initially tested in *in vitro* assays for invasion (*e.g.* modified Boyden chamber (Albini *et al.*, Cancer Res. 47:3239-3245 (1987))). This preliminary evaluation further defines the GEF that predicts activity in tumor cell invasion (as measured by the *in vitro* surrogate assays). The *in vitro* systems can also be used to
20 evaluate efficacy of combinations of "hits" with different GEF that may demonstrate activity when mixed together but not when tested alone.

E. *In vivo* evaluation

Representative compounds are preferably selected for *in vivo* evaluation
25 based upon their potency in the GEF assays. The efficacy of compounds in suppressing tumor invasion can be assessed by a number of methods, including metastatic growth of human tumor xenografts in nude, athymic mice or the invasion of tumor cells implanted on the renal capsule.

30 F. GEF definition and lead compound optimization

The GEF profile that correlates with *in vitro* and *in vivo* efficacy can be used directly as a means of optimizing "lead" compounds. This will be an essential step for any combinations of compounds that are active in the *in vitro* bioassays, since the combination therapy will be difficult to evaluate in *in vivo* assays due to probable

pharmacokinetic differences of the components of the mixture. At this stage of analysis, EC50s and maximal responses for the derivative compounds for each assay are considered.

Depending upon the selectivity requirements for the desired drug, it may be useful to incorporate additional assays into the screening panel at this stage. In that case, additional *in vivo* evaluation is necessary to validate the predictability of the new GEF for *in vivo* efficacy and selectivity.

V. Identification of Agents that Prevent or Inhibit Breast Tumor Progression

A. Background

The progression of breast cancer (BC) from a hormone-dependent, well-differentiated carcinoma to a more advanced stage lesion is marked by the loss of estrogen receptor (ER) function, decreased estrogen-cadherin (E-cadherin) expression or function, and increased vimentin expression. This progression resembles the epithelial-mesenchymal transition (EMT) (Hay, Acta Anat. 154:8-20 (1995)) that occurs during embryonic development. The advanced stage breast cancer cells adopt structural and functional characteristics of mesenchymal cells. Altered expression of intermediate filament proteins contribute to this phenotype (*e.g.*, decreased expression relative to less advanced cancer cells of some keratins and the induction of vimentin synthesis).

Additional changes include the decreased expression/function of cell junctional communication proteins (*e.g.*, E-cadherin, ZO-1), attachment factors (*e.g.*, integrins), and extracellular matrix proteins (*e.g.*, thrombospondin) as well as increased proteolytic activity (*e.g.*, stromelysin, MMPs). A significant proportion of late stage, advanced breast cancers (ABC) are represented *in vitro* by cultured BC cells that exhibit hormonal independence, decreased intercellular communication and adhesion, enhanced motility, and increased invasiveness through a reconstituted basement membrane (*i.e.*, matrigel) (Thompson *et al.*, J. Cell Physiol. 150:534-544 (1992)).

Since motile and invasive abilities are the primary distinguishing characteristics of ABC cells, we have designed experimentation to identify Gene Expression Fingerprints (GEFs) that can be substituted for the phenotypic assays generally used to measure these activities. Additional GEFs can be designed to substitute for other assays typically used to measure cancer cell progression, such as proliferation (*e.g.*, proliferative activity), apoptosis (*e.g.*, apoptotic response), angiogenesis (*e.g.*, angiogenic activity), differentiation, inflammation, and cell-cell or cell-matrix interaction.

The strategy is to identify genes whose expression is changed in the majority of ABCs and is also modulated during the process of tumorigenesis or tumor/metastasis suppression. Genes in the set of common differentially expressed genes whose expression is altered by known anti-invasive or anti-metastatic drugs will be preferentially included in a GEF used for drug screening. The GEFs will be diagnostic for ABC and predictive of drug efficacy in the treatment of ABC. The alteration of the GEF of the screening cell line(s) identifies a compound as a potential lead for further optimization.

B. Developing Diagnostic GEFs for Weakly and Highly Invasive Breast Cancer

In order to derive a GEF that can be employed in compound screens for agents that prevent progression to or inhibit the invasive and/or metastatic activity of breast tumors, we began by identifying gene expression changes that are commonly found in BC cell lines relative to normal cells. For these studies, we analyzed fourteen established cell lines derived from clinical specimens cultured from primary or metastatic samples obtained from patients diagnosed with infiltrating ductal carcinoma, which is the most prevalent type of breast cancer (Table 5, Groups I-III). Many of these cell lines have been extensively characterized for their *in vitro* growth characteristics and invasive ability as well as their *in vivo* tumorigenic and metastatic capacity. Expression of the informative marker genes ER, E-cadherin, and vimentin separates the BC cell lines into three groups [Table 5: group I is ER-positive (ER+), E-cadherin positive (E-cad+), vimentin-negative (Vim-); group II is negative for all markers; group III is negative for ER and E-cadherin expression, but positive for vimentin expression]. When categorized based upon their invasive ability in the Boyden chamber assay, these BC cell lines are separated into only two groups: a weakly invasive (Inv-w) one (encompassing cell lines in groups I and II) and a highly invasive (Inv-h) one (group III). It is noteworthy that all of the BC cell lines that express vimentin are highly invasive and exhibit a characteristic stellate morphology when cultured in matrigel. *In vivo*, the cells in this group are the only BC cell lines that are capable of forming metastases to either the lung and lymph nodes (*i.e.*, MDA231, Hs578T, MDA435) or the brain (*i.e.*, MDA435) (Price *et al.*, Cancer Res. 50:717-721 (1990)).

Table 5. Characteristics of Human Mammary Epithelial Cells Employed in this Study

| Cell Line | Specimen Origin | Tumorigenicity | Matrigel Morphol | Invasion (Boyden) | Marker Gene Expression ER E-cad Vim |
|-----------|----------------------------------|----------------|------------------|-------------------|--|
| 76N | Reduction Mammoplasty | nd | nd | nd | nd |
| 184B5 | Reduction Mammoplasty; +BP | - | fused | - | nd |
| MCF-10A | Fibrocystic Breast Disease | - | fused | - | + |
| 006FA-2B | Fibroadenoma-HPV imm | nd | fused** | nd | nd |
| HBL-100 | Milk epithelial cells | variable | stellate** | nd | nd |
| T-47D | Infiltrating Ductal Ca; PE | + | fused | + | + |
| ZR-75-1 | Infiltrating Ductal Ca; Ascites | + | fused | + | + |
| MCF-7 | Adenocarcinoma; PE | + | fused | + | + |
| BT-483 | Papillary Invasive Ductal Ca | + | fused | + | + |
| MDA-361 | Breast adenocarcinoma; brain met | + | fused | + | + |
| BT-474 | Invasive Ductal Ca | + | fused | + | + |
| BT-20 | Adenocarcinoma | + | nd | nd | nd |
| MDA-468 | Metastatic Adenocarcinoma; PE | + | fused | + | - |
| SKBR-3 | Adenocarcinoma | + | spherical | + | - |
| MDA-453 | Metastatic Ca; PE | + | spherical | + | - |
| BT-549 | Papillary Invasive Ductal Ca | - | stellate | +++ | - |
| Hs578T | Ductal Ca | +, met | stellate | +++ | - |
| MDA-231 | Adenocarcinoma | +, met | stellate | +++ | - |
| MDA-435S | Metastatic Ductal Adenocarcinoma | +, met | stellate | ++ | - |

Group I

Group II

Group III

For each cell line, the specimen source and the pathology evaluation indicated in the initial publication describing cell line

establishment are listed. BP, benzopyrene; PE, pleural effusion; Ca, carcinoma.

Tumorigenicity is recorded as + if the cell line has been reported to produce palpable tumors as xenografts in nude athymic or SCID mice. Met, metastatic cell lines. * estrogen pellets required for tumor formation. The HBL-100 cell line is given a "variable" designation since it has been shown to form tumors after extended passage in culture.

Description of the morphology of cells cultured in matrigel and their activity in the Boyden chamber invasion assay are taken from Sommers, CL et al Breast Cancer Res. & Treat. 31: 325-335 (1994). ** data from this study.

Expression of the mRNA and protein of marker genes estrogen receptor (ER) E-cadherin (E-cad) and vimentin (Vim) was obtained from literature reports.

nd, not determined

The gene expression profiles for all of the BC cell lines that represent different clinical stages and phenotypic states in BC progression have been determined by using cDNA arrays obtained from Clontech (*i.e.*, Human Atlas I). This analysis can be expanded to include additional genes (*e.g.*, other arrays, cDNA libraries) and cell sources. As a reference for these studies, we analyzed the gene expression patterns in MCF10A, a spontaneously immortalized "normal" mammary epithelial cell (MEC) line derived from a patient with fibrocystic breast disease (Soule *et al.*, Cancer Res. 50:6075-6086 (1990)). The gene expression profiles of additional "normal" cell cultures (*i.e.*, 76N MEC strain (Band and Sager, Proc. Natl. Acad. Sci. U.S.A. 86:1249-1253 (1989) and 184B5 benzopyrene-immortalized MEC (Stampfer and Bartley, Proc. Natl. Acad. Sci. U.S.A. 82:2394-2398 (1985)) derived from reduction mammoplasty specimens were also obtained. RNA from each of the cell lines was isolated and used to prepare a radiolabeled complex cDNA probe for hybridization to the Atlas I arrays. These filters contain cDNA fragments corresponding to 588 different genes that represent six functional gene classes, including oncogenes and tumor suppressor genes, genes involved in cell cycle control, cell-cell interactions, apoptosis, and signal transduction pathways. Approximately 300 of the 588 genes were detectable in these analyses indicating that over half of the genes present on the Atlas I array are expressed in human mammary epithelial cells. The hybridization signals from each cDNA spot were quantitated and compared with the signals obtained for the same gene in the arrays hybridized with a probe prepared from the reference MCF10A RNA.

An important component of the development of a GEF for compound screening is the identification of gene expression changes that can be used to discriminate between tumor-derived and "normal" cells as well as highly invasive and weakly invasive tumors. This is particularly critical in developing strategies to screen for anti-cancer drugs because cancer is the result of genomic instability and accumulated somatic mutations that lead to complex changes in gene expression. We therefore searched for genes whose expression was found to be commonly altered in tumor vs "normal" cells or in a subset of tumor cells (*e.g.*, in the four highly invasive BC cell lines). Table 6 lists the genes whose expression was frequently altered in the tumor cells relative to the reference "normal" control. The values correspond to the number of cell lines in which changes in mRNA level of at least two-fold were observed for the indicated gene. Out of the 28 genes listed, 11 were differentially expressed in the majority of the tumor cell lines compared to the reference "normal" control (Table 6). The plectin gene was

differentially expressed in all 14 BC cell lines, whereas the levels of the B-myb, transferrin R, and ICH-2 protease genes changed in 8 of the 14 cell lines (see Table 6).

Table 7A shows the fold-differences in mRNA level observed for these genes in each of the cell lines relative to its expression in the reference MCF10A. The expression of most of the genes (*i.e.*, 8/11) was decreased in the BC cells relative to "normal" cells. The other three genes (*i.e.*, B-myb, MacMarcks, and transferrin R) showed elevated expression in the BC cell lines. Other "normal" cells (*i.e.*, 76N and 184B5) exhibited minimal alteration in the expression of these genes (Figure 3A and data not shown). The pattern of expression changes (*i.e.*, increases or decreases relative to "normal" cells) for these genes represent "tumor-associated" changes found in cultured breast tumor cell lines.

We also identified genes whose expression changed primarily in BC cell lines that were categorized as either weakly or highly invasive. Table 6 delineates the number of cell lines in either the weakly or highly invasive groups that showed differential expression of the indicated genes. Two of the genes (*i.e.*, GST P and integrin A-3) were differentially expressed relative to "normal" in all 10 cell lines that have poor invasive ability; the c-jun gene was differentially expressed in all four highly invasive cell lines. The actual changes in expression level measured for each of these genes is tabulated (Table 7B). In contrast to the "tumor-associated" genes described above, most of the genes associated with either weakly or highly invasive cell lines were over-expressed in those cells relative to the "normal" cells. For the c-jun gene, all of the highly invasive cell lines express higher mRNA levels than the reference "normal". For the GST P gene, all 14 cell lines express less mRNA than the reference, but the highly invasive BC cell lines have higher levels of GST P mRNA than the weakly invasive lines, as indicated by the smaller negative value changes. These data demonstrate that some genes are differentially expressed (or repressed) in the weakly invasive cell lines. Other genes are differentially expressed (or repressed) in the more aggressive, highly invasive tumor cell lines.

Table 6.

| Gene | # of BC Lines w/ Expression Change | | |
|----------------|------------------------------------|------------|------------|
| | All Tumors | Weakly Inv | Highly Inv |
| B-myb | 8 | | |
| MacMarcks | 12 | | |
| Transferrin R | 8 | | |
| INTEGRIN A-6 | 12 | | |
| INTEGRIN B-4 | 13 | | |
| LOW-AFF NGF R | 12 | | |
| CDK inh p21 | 13 | | |
| GC-Box BP | 12 | | |
| Plectin | 14 | | |
| Alb D-box BP | 10 | | |
| ICH-2 PROTEASE | 8 | | |
| GATA-3 | | 8 | 0 |
| RABP II | | 6 | 0 |
| ERBB-3 | | 5 | 0 |
| HOX C1 PROT | | 6 | 0 |
| G NUC BP G-S | | 9 | 2 |
| ID-2 | | 6 | 0 |
| TOB | | 8 | 0 |
| INTEGRIN A-3 | | 10 | 1 |
| DB1 | | 6* | 1* |
| GST P | | 10* | 4* |
| Fra-1 | | 5* | 4* |
| c-jun | | 1 | 4 |
| bFGF R | | 0 | 3 |
| INTEGRIN A-5 | | 0 | 2 |
| N-cadherin | | 0 | 3 |
| TyrK R axl | | 0 | 3 |
| IL-8 | | 0 | 3 |
| Total Analyzed | 14 | 10 | 4 |

The number of cell lines with changes in expression of the indicated gene relative to MCF10A is provided. Only fold-changes greater than 2 were scored.

*direction or degree of expression change is different in weakly vs. highly invasive cells

Table 7A.

| Gene | GB# | GENE EXPRESSION CHANGES IN BREAST CANCER CELL LINES* | | | | | | | | | |
|----------------|--------|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | WEAKLY INVASIVE | | | | | HIGHLY INVASIVE | | | | |
| | | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w |
| B-myb | X13293 | BT483 | T47D | ZR75-1 | MCF7 | MDA361 | BT474 | MDA468 | BT20 | SKBR3 | MDA453 |
| MacMarcks | X70326 | 23 | 5 | 6 | 8 | 8 | 12 | 7 | 9 | 4 | 3 |
| Transferrin R | X01060 | | 6 | | 2 | 3 | 3 | | 5 | 8 | 9 |
| INTEGRIN A-6 | X53586 | -13 | -13 | -18 | -8 | -13 | -6 | -7 | -4 | -40 | -7 |
| INTEGRIN B-4 | X53587 | -57 | -6 | -9 | -4 | -5 | -7 | -9 | -8 | -12 | -7 |
| LOW-AFF NGF R | M14764 | -3 | -4 | -3 | | -6 | -5 | -3 | -8 | -6 | -7 |
| CDK Inh p21 | U09579 | -2 | -6 | -4 | | -6 | -13 | -4 | -12 | -12 | -10 |
| GC-Box BP | D14520 | -17 | -34 | -6 | -7 | -3 | -4 | -4 | -10 | -7 | -12 |
| pledin | M63618 | -49 | -11 | -8 | -8 | -28 | -12 | -4 | -10 | -7 | -12 |
| Ab D-box BP | D28468 | -3 | -2 | | | | -4 | -3 | -3 | -4 | -3 |
| ICH-2 PROTEASE | U28014 | -7 | -4 | | -15 | | -3 | 3 | 4 | 4 | 4 |

Table 7B.

| Gene | GB# | GENE EXPRESSION CHANGES IN BREAST CANCER CELL LINES* | | | | | | | | | |
|----------------|--------|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | | WEAKLY INVASIVE | | | | | HIGHLY INVASIVE | | | | |
| | | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w | ER+/E-cad+/Vim-/Inv-w | ER-/E-cad+/Vim-/Inv-w |
| GATA-3 | X55122 | BT483 | T47D | ZR75-1 | MCF7 | MDA361 | BT474 | MDA468 | BT20 | SKBR3 | MDA453 |
| RABP II | M68867 | 58 | 24 | 10 | 73 | 10 | 15 | 29 | 4 | 4 | 21 |
| ERBB-3 | M29366 | 14 | 7 | | 13 | | 12 | | | | 24 |
| HOX C1 PROTEIN | M16937 | 4 | 3 | 3 | 4 | 3 | 6 | | 4 | 4 | 10 |
| GN BP G-S | M14631 | 4 | 3 | 4 | 4 | 4 | 4 | | 4 | 7 | 3 |
| ID-2 | M97796 | 4 | 5 | 5 | 7 | 11 | 7 | | 2 | 16 | 5 |
| TOB | D38305 | 4 | 5 | 4 | 3 | 10 | 4 | | 2 | 3 | 3 |
| INTEGRIN A-3 | M59911 | -13 | -10 | -8 | -17 | -4 | -3 | -4 | -2 | -14 | -5 |
| DB1 | D28118 | | -4 | -4 | | -2 | | -5 | -4 | -4 | -4 |
| GST P | X15480 | -2007 | -359 | -501 | -886 | -327 | -447 | -8 | -303 | -455 | -585 |
| Fra-1 | X16707 | -3 | | -3 | -5 | | -4 | | 4 | -5 | |
| c-Jun | J04111 | | | | | | | | | | |
| bFGF R | M37722 | | | | | | | | | | |
| INTEGRIN A-5 | X06256 | | | | | | | | | | |
| N-cadherin | M34084 | | | | | | | | | | |
| TyrK R axl | M76125 | | | | | | | | | | |
| IL-8 | Y00787 | | | | | | | | | | |

* Change in expression in BC cell lines relative to MCF10A

The consensus GEFs for weakly and highly invasive cancers are graphically depicted in Figure 3A. The GEF of a normal MEC strain (*i.e.*, 76N) is also shown for comparison. Three sub-profiles can be distinguished: a tumor-associated GEF comprising 11 genes (Figure 3A, left-handed striped bars (bars having a stripe angling downward from left to right)), a GEF representative of weakly invasive carcinomas comprising 8 genes (Figure 3A, solid bars), and a GEF diagnostic for highly invasive, ABC comprising 6 genes (Figure 3A, right-handed striped bars (bars having a stripe angling upward from left to right)). Three genes show distinguishable differential expression patterns in both weakly and highly invasive cell lines relative to "normal" (Figure 3A, stippled bars) and are therefore diagnostic for either invasive state. These data strongly suggest that the expression pattern of the 28 genes in an uncharacterized cell line could be used as a means of predicting its tumorigenic and invasive potential. We analyzed the GEFs of two cell lines that have not been tested for invasive activity. One of these is a cell line derived in our laboratory from a breast fibroadenoma tissue specimen that was cultured and immortalized by transfection with the HPV E6/E7 oncogenes. The other is the HBL100 cell line that was established from human milk epithelial cells and subsequently shown to contain integrated SV40 genomic sequences that encode the T antigen protein (Vanhamme and Szpire, Carcinogenesis 9:653-655 (1988)). The expression profiles of these two cell lines are shown in Figure 3B. From these patterns, we predict that the HBL-100 cell line is a tumor-derived mesenchymal-like, highly invasive cell line; in contrast, the 006FA-2B cells are significantly different from "normal" immortal HMEC such as the MCF10A and 184B5, but do not exhibit the differential gene expression pattern of either of the tumor cell phenotypes profiled in these studies. The growth characteristics in matrigel of these two cell lines were assayed in order to determine whether they demonstrated the morphology associated with the phenotypes predicted by their GEF. In agreement with the GEFs for these cell lines, the 006FA-2B adopted a fused morphology in matrigel whereas the HBL-100 grew with the stellate morphology characteristic of mesenchymal cells with highly invasive ability (data not shown).

The GEFs identified in cell culture models of breast cancer have value in staging clinical specimens or evaluating responses to drug therapy. The gene expression patterns were determined for three tumor biopsies obtained from patients with moderately differentiated infiltrating ductal carcinomas of the breast and compared with the gene expression profile of normal breast tissue. In the profiles shown in Figure 3C, the

characteristic tumor-associated GEF is found in all three of the tumors, being most pronounced in tumors T8911044 and T8911045. Furthermore, all of these tumors exhibit a GEF that is correlated with weakly invasive tumors. These data indicate that GEFs similar to those described here useful in the diagnosis and treatment of cancer patients. They also suggest that the cultured cells faithfully reproduce some of the gene expression changes observed in the *in vivo* tumor environment.

C. Development of Process-Associated GEFs

The GEFs identified up to this point are diagnostic of the phenotypic states of highly and weakly invasive cells. These gene expression differences are valuable in diagnostic applications. Also of interest is whether gene expression differences are able or sufficient to report the activity of anti-invasive or metastatic drugs. The selection of a subset of these 28 genes that is most useful in predicting drug efficacy is assisted by determining whether any of these genes are associated with the process of malignant progression. To that end, we measured gene expression changes that occur during cellular transformation as well as tumor and/or metastasis suppression. Models for these processes include oncogene-transformed normal HMEC, tumor suppressor gene-transfected tumor cells, and treatments with anti-neoplastic drugs or differentiating agents. These studies can include analysis of gene expression patterns following treatment of cells *in vivo* or *in vitro* under a variety of conditions, including, but not limited to, culture on matrigel, on low attachment tissue culture plates, or with other cell types. Knowledge of the gene expression changes that occur during the conversion of a weakly or non-invasive BC cell to one with highly invasive activity by treatment with growth factors (*e.g.*, EGF, scatter factor) or transfection with oncogenes (*e.g.*, v-ras) are particularly valuable. Additional model systems that recapitulate the EMT (*e.g.*, treatment with anti-E-cadherin antibodies) can also be employed to define the genes that report the invasive properties of BC cells. Information concerning gene expression changes that correlate with the reduction in invasive capacity in response to treatment with drugs or invasion-suppressor gene products is also desirable for deriving the GEF for compound screening.

Normal limited lifespan HMEC can be immortalized by expression of the SV40 T antigen, the HPV E6 oncogene, or selected p53 mutant proteins (Band, Intl. J. Oncol. 12:499-507 (1998)). Using the Atlas I array, we measured the gene expression changes that occurred in HMEC immortalized by infection with mutant p53-expressing

retrovirus (Gao *et al.*, Cancer Res. 56:3129-3133 (1996)). The expression level of 13 genes was affected following immortalization with three different p53 mutant proteins that act as dominant-negative inhibitors of p53 function; notably, 6 of them are included in the "tumor-associated" GEF (Table 8). These data suggest that inactivation of p53 is a critical determinant of the decreased gene expression observed for those genes. These data also imply that these genes are reporters of a critical step in the process of tumorigenesis — that of cellular immortalization. They also infer that p53 inactivation is important in the generation of tumors represented by many of these BC cell lines. Mutation of p53 is an event that is associated with the majority of breast carcinomas. It is of interest that the tumor biopsies also showed decreased expression of 4 of these 6 genes relative to normal tissue controls (Table 8). These studies demonstrate a means of identifying a GEF that is representative of the process of tumor formation. The genes comprising that GEF which are also identified as diagnostic for ABC would be included in the gene-cell combinations used in the drug screen.

The identification of genes that predict anti-invasive drug activity is aided by measuring the gene expression changes resulting from treatment of highly invasive cells with anti-invasive or anti-metastatic drugs. By comparing the effects of anti-invasive compounds that have different known mechanisms of action, a common set of genes whose expression changes report anti-invasive activity can be derived. Also important is the determination of the gene expression changes caused by drugs that are ineffective in blocking invasion, but have other anti-neoplastic properties (*e.g.*, pro-apoptotic, anti-angiogenic, anti-proliferative), as well as compounds that are modulators of signaling pathways that do not result in the inhibition of invasion. In the studies presented here, we tested taxol, mevastatin, sodium butyrate, retinoic acid (RA), and caffeic acid (CA). Taxol's efficacy is reported to be dependent upon its inhibition of microtubule formation, while mevastatin inhibits HMG CoA reductase and indirectly protein prenylation, thereby leading to cell cycle arrest in the G1 phase. Sodium butyrate is a differentiating agent that causes histone acetylation and transcriptional activation. RA has anti-proliferative and differentiating effects in some BC cell lines (*i.e.*, ER+), but is ineffective in others (*i.e.*, ER-negative). Both taxol and mevastatin are capable of blocking the development of the characteristic stellate mesenchymal cell morphology of MDA231 cells, while sodium butyrate is not effective (data not shown). Taxol has also been shown to prevent invasion of MDA231 in the Boyden chamber assay (Sasaki and Passanti, Biotechniques 24:1038-1043 (1998)) and mevastatin inhibits

mammary tumor metastases *in vivo* (Alonso *et al.*, Breast Cancer Res. Treat. 50:83-93 (1998)). The highly invasive MDA231 BC cells were treated with these compounds under conditions (*i.e.*, concentration and time) reported to have maximal effects with little toxicity. Taxol, mevastatin, and butyrate treatment caused changes of greater than two-fold in the expression of approximately 10% of the expressed Atlas I array genes (*i.e.*, taxol: 27/300; mevastatin: 33/300; butyrate: 39/300), while little effect was observed with either RA or CA treatment. The gene expression profiles of each of these compounds are readily distinguishable from each other (Figure 4). Significantly, 12 of the 28 genes identified as potential reporters of either tumorigenicity or stage of invasiveness are modulated by one or more of these drugs. Moreover, the direction of the gene expression change elicited by these drugs for 11 of these 12 genes is towards a more "normal" or less invasive GEF (Table 8). For example, the expression of 7 genes that were either repressed or enhanced in the highly invasive MDA231 cancer cells relative to "normal" were reversed. The expression changes for four of the genes (*i.e.*, RABP II, Integrin A-3, DB1, and GST P) are in the direction towards a less invasive GEF (*e.g.*, RABP II expression is elevated following drug treatment to levels that are higher than the "normal" cells similar to the expression change in weakly invasive cell lines). Such data suggest that these genes are reporters of drug activities that affect malignant progression, but they do not necessarily identify genes that can be used to predict anti-invasive efficacy *per se*. The subset of genes that is commonly regulated by both mevastatin and taxol, but not butyrate (*i.e.*, GC-Box BP, RABP II, DB1), is likely to report anti-invasive effects, since both of these agents are presumed to have anti-invasive activity based upon matrigel morphology studies while butyrate does not. Evaluation of additional drug treatments that have anti-invasive effects as well as those with only anti-proliferative or pro-apoptotic effects enables further fine-tuning of the GEF that is most predictive of drug efficacy, selectivity for invasive action, and potential toxicity.

Tabl 8.

GENE EXPRESSION CHANGES

| | | Diagnostic | | | Process | | | |
|------------------|---------------|------------------|------------|-------------|------------------------------|------------------|-----|------|
| | | in BC Cell Lines | | in Biopsies | Tumorigenesis p53 inactiv | Anti-cancer Drug | | |
| | | Weakly Inv | Highly Inv | | | Tax | Mev | Buty |
| Tum r-ass clated | Gene | | | | | | | |
| | B-myb | + | + | | | | | |
| | MacMarcks | + | + | | | | | |
| | Transferrin R | + | + | | | | | |
| | INTEGRIN A-6 | - | - | - | - | | | + |
| | INTEGRIN B-4 | - | - | | - | | | |
| | LOW-AFF NGF R | - | - | - | - | | + | |
| | CDK Inh p21 | - | - | - | - | | + | |
| | GC-Box BP | - | - | - | | + | + | |
| | plectin | - | - | - | - | | | |
| | Alb D-box BP | - | - | | - | | + | |
| ICH-2 PROTEASE | - | - | | | | | | |
| Weakly Invasive | GATA-3 | + | | + | | | | |
| | RABP II | + | | + | | + | + | |
| | ERBB-3 | + | | + | | | | |
| | HOX C1 | + | | | | | | |
| | GN BP G-S | + | | | | | | |
| | ID-2 | + | | + | | | | |
| | TOB | + | | + | | | | |
| | INTEGRIN A-3 | - | | | - | | | |
| Highly Invasive | DB1 | - | | | | - | - | |
| | GST P | - - - | - | | | | - | |
| | Fra-1 | - | + | - | | | | - |
| | c-jun | | + | | | | | |
| | bFGF R | | + | | | | | |
| | INTEGRIN A- 5 | | + | | | | | |
| | N-cadherin | | + | | | | | |
| | TyrK R axd | | + | | | - | - | - |
| | IL-8 | | + | | | + | + | + |

The direction of expression change for each of the indicated genes is tabulated under the Diagnostic heading for differences in BC Cell Lines and Tumor Biopsies relative to MCF10A and normal breast tissue, respectively (data from Tables 7A and 7B and Fig. 3C). Under the Process heading, genes modulated in cells immortalized by p53 inactivation relative to their limited lifespan counterparts are indicated in the Tumorigenesis column. The direction of gene expression change in the highly invasive MDA231 cells in response to treatment with either taxol (taxol), mevastatin (mev), or sodium butyrate (buty) is provided in the Anti-cancer drug column.

D. Defining a GEF for Anti-Invasive Drug Screening

The studies described here have derived a GEF incorporating the expression of 28 genes that is useful in distinguishing between weakly and highly invasive BC cell lines and tumor biopsies. Within the GEF there is a subset of gene expression changes associated with all BC cell lines and tumors (*i.e.*, tumor-associated GEF). In combination with the tumor-associated GEF, two other distinct sub-GEFs define weakly vs. highly invasive cancers. Experiments using tumor progression model systems (*i.e.*, p53 inactivation) and anti-neoplastic drug treatments have identified genes within the 28 that are modulated in the process of tumorigenesis or during the inhibition of invasion.

The precise GEF that predicts anti-invasive drug efficacy is a change in the expression of a subset of the 28-gene GEF representative of highly invasive cancer cells. That subset is determined by a selection procedure similar to the one used to derive the diagnostic GEFs. Genes commonly affected by drugs or other agents which modulate the invasive phenotype are compared with the diagnostic GEF to derive the common gene expression changes; this produces a GEF predictive of drug efficacy. The gene-cell combinations used to create the screen for anti-invasive compounds includes the highly invasive MDA231 cell line and at least two genes from each of the sub-GEFs described above (*i.e.*, tumor-associated, weakly invasive, and highly invasive). Gene and cell line selection also considers data from drug treatment of the other highly invasive cell lines as well as weakly invasive ones. The GEF screen can be carried out in more than one cell line either in mixed or parallel cultures.

E. Materials & Methods

1. Cell Culture and Compound Treatment

The 76N human MEC strain and the 184B5 benzopyrene-immortalized human MEC line were cultured in DFCI-1 medium (Band and Sager, Proc. Natl. Acad. Sci. U.S.A. 86:1249-1253 (1989)). The 006FA-2B cell line was established from a benign fibroadenoma tissue sample by co-transfecting the cultured organoids with plasmid vectors encoding the HPV16 E6 and E7 oncogenes and a selectable SVneo plasmid using a standard calcium phosphate-mediated procedure. 006FA-2B is one of several stable epithelial cell clones with extended lifespan that were selected using G418 (100 μ g/ml, Gibco). MCF10A, HBL-100, T47D, ZR75-1, MCF7, BT483, MDA361, BT474, BT20, MDA468, SKBR3, MDA453, BT549, Hs578T, MDA231, and MDA435S

cells were obtained from the ATCC (Rockville, MD) and initially cultured in the ATCC-recommended medium. To determine the steady state gene expression profiles of the breast tumor lines, the cells were cultured to 80-90% confluency in α -MEM medium [alpha-modified MEM supplemented with 1 mM HEPES, 2 mM glutamine, 0.1 mM MEM non-essential amino acids, 1.0 mM sodium pyruvate, 50 μ g/ml gentamicin, 1.0 μ g/ml insulin (all from Gibco, Gaithersburg, MD), and 10 % FBS (Intergen)]. To evaluate the effect of selected compounds on gene expression in the MDA231 cell line, cells were plated (10^6 /100 mm dish) in α -MEM medium and allowed to attach overnight. Cells were fed with fresh medium containing 3 mM sodium butyrate (Specialty Media, Inc. Lavallette, NJ), 5.0 μ M taxol (Molecular Probes, Inc., Eugene, OR), 10^{-8} M caffeic acid, 1.0 M retinoic acid, or 20 μ M mevastatin (all from Sigma) and cell monolayers harvested 72 hours (h) later for RNA isolation.

2. Gene Expression Analysis

Total RNA from cell lines and compound-treated cells was isolated by the guanidinium-isothiocyanate-CsCl gradient procedure (Chirgwin *et al.*, Biochemistry 18: 5294-5299 (1979)). Total RNA from normal and tumor tissue specimens was obtained from BioChain Institute, Inc (San Leandro, CA).

The preparation of radioactively labeled cDNA from total RNA (5 μ g) was performed essentially as described in the Clontech Atlas I cDNA array hybridization kit protocol. The only exceptions were the step for removal of unincorporated nucleotide triphosphate, which was carried out using a G50 spin column and the length of prehybridization, which was increased to at least 6 h. The probe concentration routinely employed in the hybridization reactions was $0.7-1.0 \times 10^6$ counts per minute/milliliter (cpm/ml).

3. Image Analysis of Clontech Atlas I cDNA Expression Arrays

The probe intensities at each target (cDNA) spot on the Atlas I arrays were quantitated using the "Array Vision" software package from Imaging Research, Inc. (St. Catherine, Ontario, Canada). The grid definition protocol was used in this analysis with an automated algorithm to finely adjust the grid to overlay the targets. Each target in the array was scanned using the Storm Phosphorimaging System by Molecular Dynamics, Inc. (Sunnyvale, CA) and a data table was constructed of the average PSL x area values (the PSL value per pixel times the area in mm of the target) corrected for

background and reference normalization. An average background was determined from a selected blank region of the array and a reference value for normalization was generated using the average of the signals of all of the targets on the array. The ratios and z-score differences between two samples are calculated and differentially expressed genes are identified from a common set of thresholded ratios and differences. For these analyses, ratio thresholds were 2-fold and z score values were 0.3.

All references cited herein are expressly incorporated by reference in their entirety for all purposes.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modification may be practiced within the scope of the appended claims.

We claim:

1. A method for grouping test compounds into classes, the method comprising:

- (a) exposing a cell culture or cultures comprising at least two gene-cell combinations to a test compound to generate an exposed cell culture or cultures, wherein each of the at least two gene-cell combinations comprises a unique combination of a particular gene and a cell of a particular cell type;
- (b) preparing RNA from the exposed cell culture(s);
- (c) screening RNA from (b) for mRNA of each particular gene of each of the at least two gene-cell combinations of (a) to generate a gene expression fingerprint (GEF) for the test compound;
- (d) repeating (a) - (c) for each test compound to be grouped into classes; and
- (e) comparing the GEF for each test compound tested in (a) - (d), wherein the test compounds are grouped into at least two classes based on differences or similarities in their GEFs.

2. The method of claim 1, wherein the at least two gene-cell combinations comprises at least two different genes.

3. The method of claim 1, wherein the at least two gene-cell combinations comprises at least two different cell types.

4. The method of claim 1, wherein the screening comprises PCR amplification using oligonucleotide primers specific for each gene.

5. The method of claim 1, wherein the RNA is optionally reverse transcribed into cDNA.

6. The method of claim 1 or 5, wherein the screening comprises hybridization of nucleic acid sequences specific for each gene to the RNA or cDNA.

1 7. The method of claim 1, wherein at least one gene in the at least
2 two gene-cell combinations comprises an endogenous gene under control of its native
3 promoter.

1 8. The method of claim 1, wherein at least one gene in the at least
2 two gene-cell combinations comprises a heterologous gene under control of a
3 heterologous promoter.

1 9. The method of claim 1, wherein at least one gene in the at least
2 two gene-cell combinations further comprises an internal negative control gene, wherein
3 an effect on a level of mRNA of the negative control gene in response to the test
4 compound is indicative of a toxic effect of the test compound.

1 10. The method of claim 1, wherein at least one gene in the at least
2 two gene-cell combinations further comprises an internal negative control gene, wherein
3 an effect on a level of mRNA of the negative control gene in response to the test
4 compound is indicative of a non-specific effect of the test compound.

1 11. The method of claim 1, wherein the screening further comprises
2 quantitating an effect on a level of mRNA of at least one gene in the at least two gene-
3 cell combinations.

1 12. The method of claim 1, wherein the method further comprises
2 administering a combination of two or more test compounds to the cell cultures in (a),
3 wherein a GEF is generated for the combination of said two or more test compounds.

1 13. The method of claim 1, wherein the test compound is a mimetic of
2 estrogen, p53, IFN β , TNF α , endothelin, tamoxifen, raloxifene, IFN α , IFN γ , or an anti-
3 Ha-ras ribozyme.

1 14. The method of claim 1, wherein the test compound is a peptide,
2 peptidomimetic, polypeptide, protein, ribozyme, nucleic acid, oligonucleotide, organic or
3 inorganic compound, or an animal, plant, or microbial extract.

1 15. The method of claim 1, wherein the method further comprises
2 testing a representative test compound in each class for an activity of interest *in vivo*.

1 16. The method of claim 15, wherein the representative test compound
2 is a mimetic of p53, estrogen, raloxifene, tamoxifen, or IFN β .

1 17. The method of claim 15, wherein the activity of interest is tumor
2 suppression.

1 18. The method of claim 15, wherein the activity of interest is
2 decreased bone loss.

1 19. The method of claim 15, wherein the activity of interest is anti-
2 metastatic activity, prevention of atherosclerotic lesion progression, decreased
3 inflammation in rheumatoid arthritis, improved cognitive function, or prevention of hot
4 flushes.

1 20. A method for grouping test compounds into classes, the method
2 comprising:

3 (a) exposing a cell culture or cultures comprising at least two
4 gene-cell combinations to a test compound to generate an exposed cell culture or
5 cultures, each of the at least two gene-cell combinations comprising a unique combination
6 of a particular gene and a cell of a particular cell type, wherein at least one gene in the
7 at least two gene-cell combinations is differentially expressed in first and second
8 reference states;

9 (b) preparing RNA from the exposed cell culture(s) of (a);

10 (c) screening RNA from (b) for mRNA of each particular gene
11 in each of the at least two gene-cell combinations of (a) to generate a gene expression
12 fingerprint (GEF) for the test compound;

13 (d) repeating (a) - (c) for each test compound to be grouped
14 into classes; and

15 (e) comparing the GEF for each test compound tested in (a) -
16 (d);

17 wherein the test compounds are grouped into at least two classes based on
18 differences in their GEFs.

1 (21.) The method of claim 20, wherein at least one of the first and
2 second reference states is a disease state.

1 22. The method of claim 21, wherein the disease state is cancer.

1 23. The method of claim 20, wherein the screening comprises PCR
2 amplification using oligonucleotide primers specific for each gene in the at least two
3 gene-cell combinations.

1 24. The method of claim 20, wherein the RNA is optionally reverse
2 transcribed into cDNA.

1 25. The method of claim 20 or 24, wherein the screening comprises
2 hybridization of nucleic acid probes specific for each gene in the at least two gene-cell
3 combinations to the RNA or cDNA.

1 26. The method of claim 20, wherein at least one gene in the at least
2 two gene-cell combinations further comprises an internal negative control gene, wherein
3 an effect on the mRNA level of the negative control gene in response to the test
4 compound is indicative of a toxic effect of the test compound.

1 27. The method of claim 20, wherein at least one gene in the at least
2 two gene-cell combinations further comprises an internal negative control gene, wherein
3 an effect on the mRNA level of the negative control gene in response to the test
4 compound is indicative of a non-specific effect of the test compound.

1 28. The method of claim 20, wherein the screening further comprises
2 quantitating the level of the mRNA of each gene in the at least two gene-cell
3 combinations.

1 29. The method of claim 20, wherein the method further comprises
2 testing a representative test compound in each class for a desired activity *in vivo*.

1 30. The method of claim 20, wherein the method further comprises
2 administering a combination of two or more test compounds to the cell culture(s) in (a),
3 wherein a GEF is generated for the combination of said two or more test compounds.

1 31. A method of generating a reference gene expression fingerprint
2 (GEF) for at least one reference compound for use in grouping test compounds into
3 classes, said method comprising:

4 (a) identifying at least two gene-cell combinations, each of said
5 at least two gene-cell combinations comprising a unique combination of a particular gene
6 and a cell of a particular cell type, wherein a first gene-cell combination is identified by:

7 (i) exposing host cells *in vivo* or a host cell culture of a
8 first cell type to a first reference compound;

9 (ii) preparing RNA from the exposed host cells *in vivo* or
10 the host cell culture of (ii);

11 (iii) comparing the RNA of (ii) to RNA prepared from host
12 cells *in vivo* or a host cell culture of the first cell type not exposed to the first reference
13 compound, wherein a change in a level of mRNA for a gene in cells of the first cell type
14 in response to the first reference compound identifies the gene and cells of the first cell
15 type as the first gene-cell combination for use in grouping test compounds into classes;
16 and wherein a second gene-cell combination is identified by:

17 (iv) exposing host cells *in vivo* or a host cell culture of the
18 first cell type or a second cell type to the first reference compound;

19 (v) preparing RNA from the exposed host cells *in vivo* or
20 the host cell culture of (iv);

21 (vi) comparing the RNA of (v) to RNA prepared from host
22 cells *in vivo* or a host cell culture of the same cell type as in (iv) not exposed to the first
23 reference compound, wherein a gene having an mRNA level changed in response to the
24 first reference compound is identified as a gene for use in the second gene-cell
25 combination for use in grouping test compounds into classes, said second gene-cell
26 combination being different from said first gene-cell combination and comprising the
27 identified gene and cells of the same cell type as in (iv); and

28 (b) screening RNA of (ii) and (v) for mRNA for each gene in
29 each of the at least two gene-cell combinations to generate a reference GEF for the first
30 reference compound for use in grouping test compounds into classes.

1 32. The method of claim 31, wherein the at least two gene-cell
2 combinations comprises at least two different genes.

1 33. The method of claim 31, wherein the at least two gene-cell
2 combinations comprises at least two different cell types.

1 34. The method of claim 31, wherein the screening comprises PCR
2 amplification using oligonucleotide primers specific for each gene of each of the at least
3 two gene-cell combinations.

1 35. The method of claim 31, wherein the RNA is optionally reverse
2 transcribed into cDNA.

1 36. The method of claim 31 or 35, wherein the screening comprises
2 hybridization of nucleic acid sequences specific for each gene of each of the at least two
3 gene-cell combinations to the RNA or cDNA.

1 37. The method of claim 31, wherein at least one gene in the at least
2 two gene-cell combinations comprises an endogenous gene under control of its native
3 promoter.

1 38. The method of claim 31, wherein at least one gene in the at least
2 two gene-cell combinations comprises a heterologous gene under control of a
3 heterologous promoter.

1 39. The method of claim 31, wherein at least one gene in the at least
2 two gene-cell combinations further comprises an internal negative control gene, wherein
3 an effect on a level of mRNA of the negative control gene in response to the test
4 compound is indicative of a toxic effect of the test compound.

1 40. The method of claim 31, wherein at least one gene in the at least
2 two gene-cell combinations further comprises an internal negative control gene, wherein
3 an effect on a level of mRNA of the negative control gene in response to the test
4 compound is indicative of a non-specific effect of the test compound.

1 41. The method of claim 31, wherein the screening further comprises
2 quantitating an effect on a level of mRNA of at least one gene in the at least two gene-
3 cell combinations.

1 42. The method of claim 31, wherein the first reference compound is
2 estrogen, p53, IFN β , TNF α , endothelin, tamoxifen, raloxifene, IFN α , IFN γ , or an anti-
3 Ha-ras ribozyme.

1 43. The method of claim 31, wherein the first reference compound is
2 a peptide, peptidomimetic, polypeptide, protein, ribozyme, nucleic acid, oligonucleotide,
3 organic or inorganic compound, or an animal, plant, or microbial extract.

1 44. The method of claim 31, wherein (a) - (b) is repeated for a second
2 reference compound, whereby a gene having an mRNA level changed in response to the
3 first reference compound but not the second reference compound is identified as having a
4 response specific for the first reference compound.

1 45. The method of claim 44, wherein the second reference compound
2 is different from the first reference compound and comprises a mimetic of estrogen, p53,
3 IFN β , TNF α , endothelin, tamoxifen, raloxifene, IFN α , IFN γ or an anti-Ha-ras
4 ribozyme.

1 46. The method of claim 44, wherein the second reference compound
2 is the product of a gene expressed in the host cell.

1 47. The method of claim 31, wherein the first reference compound is
2 the product of a gene expressed in the host cell.

1 48. The method of claim 31, wherein the gene is a p53 gene.

1 49. A method for grouping test compounds into classes, said method
2 comprising:

3 (a) generating a reference GEF for a reference compound
4 according to the method of claim 31;

5 (b) generating a GEF for each test compound to be grouped
6 into classes by:

7 (i) exposing a cell culture or cultures comprising the at
8 least two gene-cell combinations identified in claim 31 to a test compound to generate an
9 exposed cell culture or cultures;

10 (ii) preparing RNA from the exposed cell culture or
11 cultures of (i);

12 (iii) screening RNA of (ii) for mRNA of each gene in
13 each of the at least two gene-cell combinations of (i) to generate a GEF for the test
14 compound;

15 (iv) repeating (i) - (iii) for each test compound to be
16 grouped into classes to generate a GEF for each said test compound; and

17 (c) comparing the GEF for each test compound generated in
18 (b) with the reference GEF of (a), wherein the test compounds are grouped into at least
19 two classes based on differences or similarities between their GEFs and the reference
20 GEF.

1 50. The method of claim 49, wherein the method further comprises
2 administering a combination of two or more test compounds to the cell cultures in (a),
3 wherein a GEF is generated for the combination of said two or more test compounds.

1 51. The method of claim 49, wherein the test compound is a mimetic
2 of estrogen, p53, IFN β , TNF α , endothelin, tamoxifen, raloxifene, IFN α , IFN γ , or an
3 anti-Ha-ras ribozyme.

1 52. The method of claim 49, wherein the test compound is a peptide,
2 peptidomimetic, polypeptide, protein, ribozyme, nucleic acid, oligonucleotide, organic or
3 inorganic compound, or an animal, plant, or microbial extract.

1 53. The method of claim 49, wherein the method further comprises
2 testing a representative test compound in each class for an activity of interest *in vivo*.

1 54. The method of claim 53, wherein the representative test compound
2 is a mimetic of p53, estrogen, raloxifene, tamoxifen, or IFN β .

1 55. The method of claim 53, wherein the activity of interest is tumor
2 suppression.

1 56. The method of claim 53, wherein the activity of interest is
2 decreased bone loss.

1 57. The method of claim 53, wherein the activity of interest is anti-
2 metastatic activity, prevention of atherosclerotic lesion progression, decreased
3 inflammation in rheumatoid arthritis, improved cognitive function, or prevention of hot
4 flushes.

1 58. A method of claim 20, wherein at least one of the first and second
2 reference states comprises a change in a cellular phenotype.

1 59. The method of claim 58, wherein the change in the cellular
2 phenotype comprises a change in cellular invasiveness, apoptotic response, angiogenic
3 activity, proliferative activity, inflammation, cell-cell interaction, or cell-matrix
4 interaction.

1/6

A.



B.

**FIG. 1.**

2/6

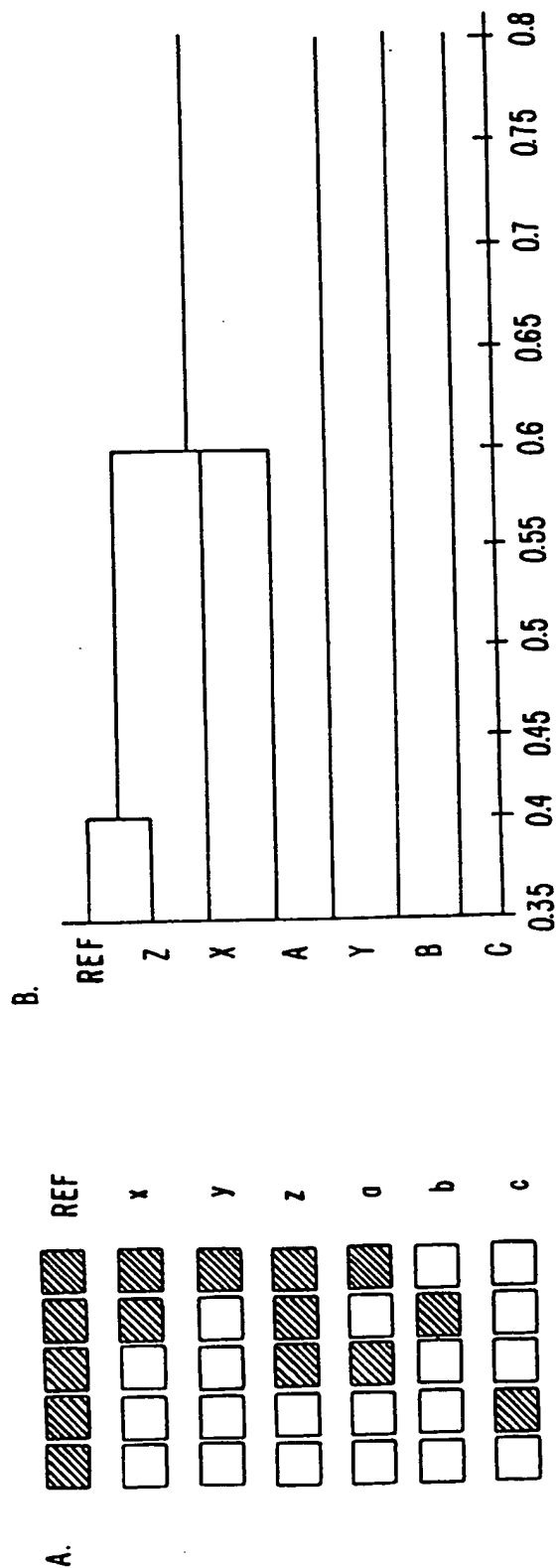


FIG. 2.

3/6

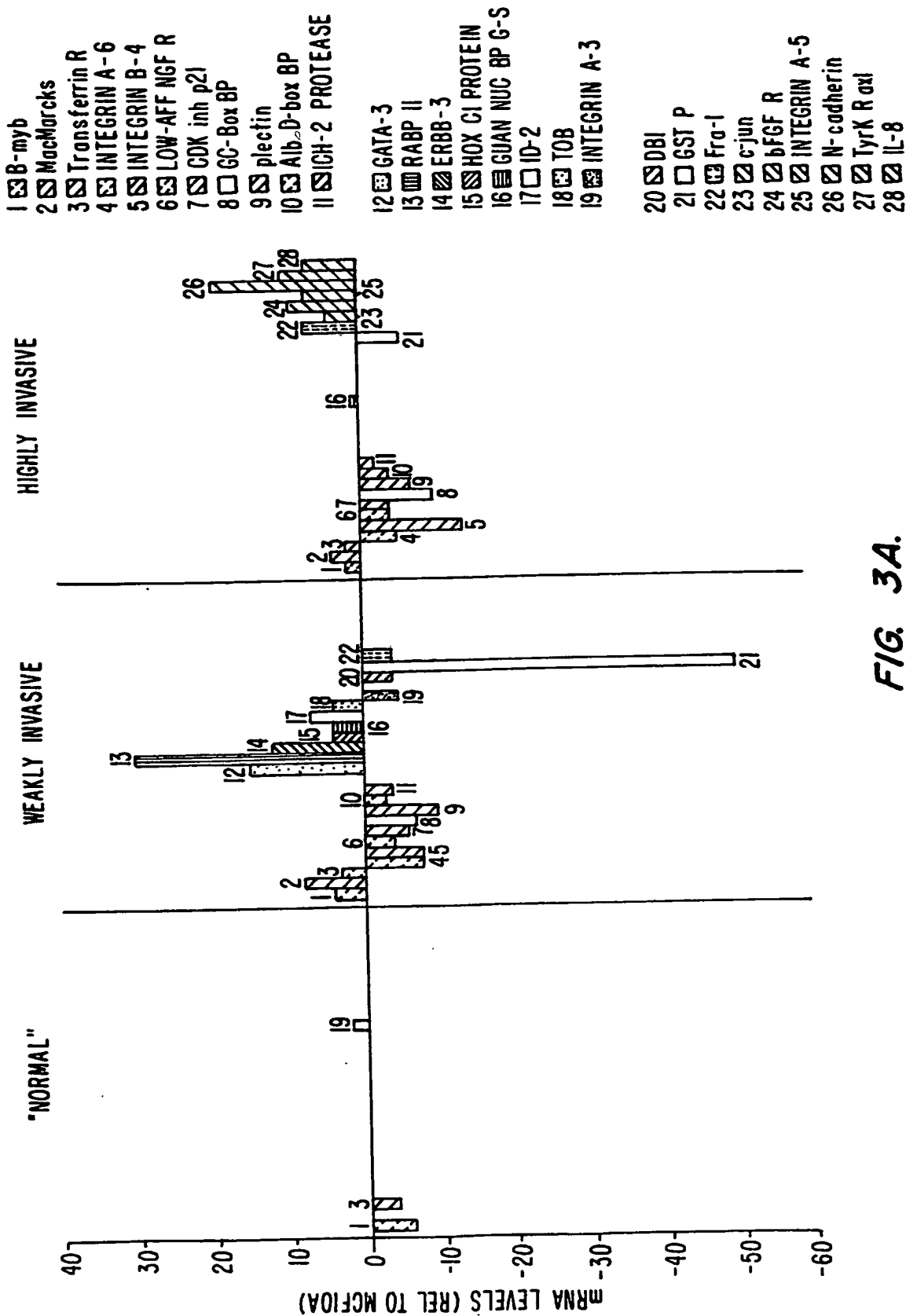


FIG. 3A.

4/6

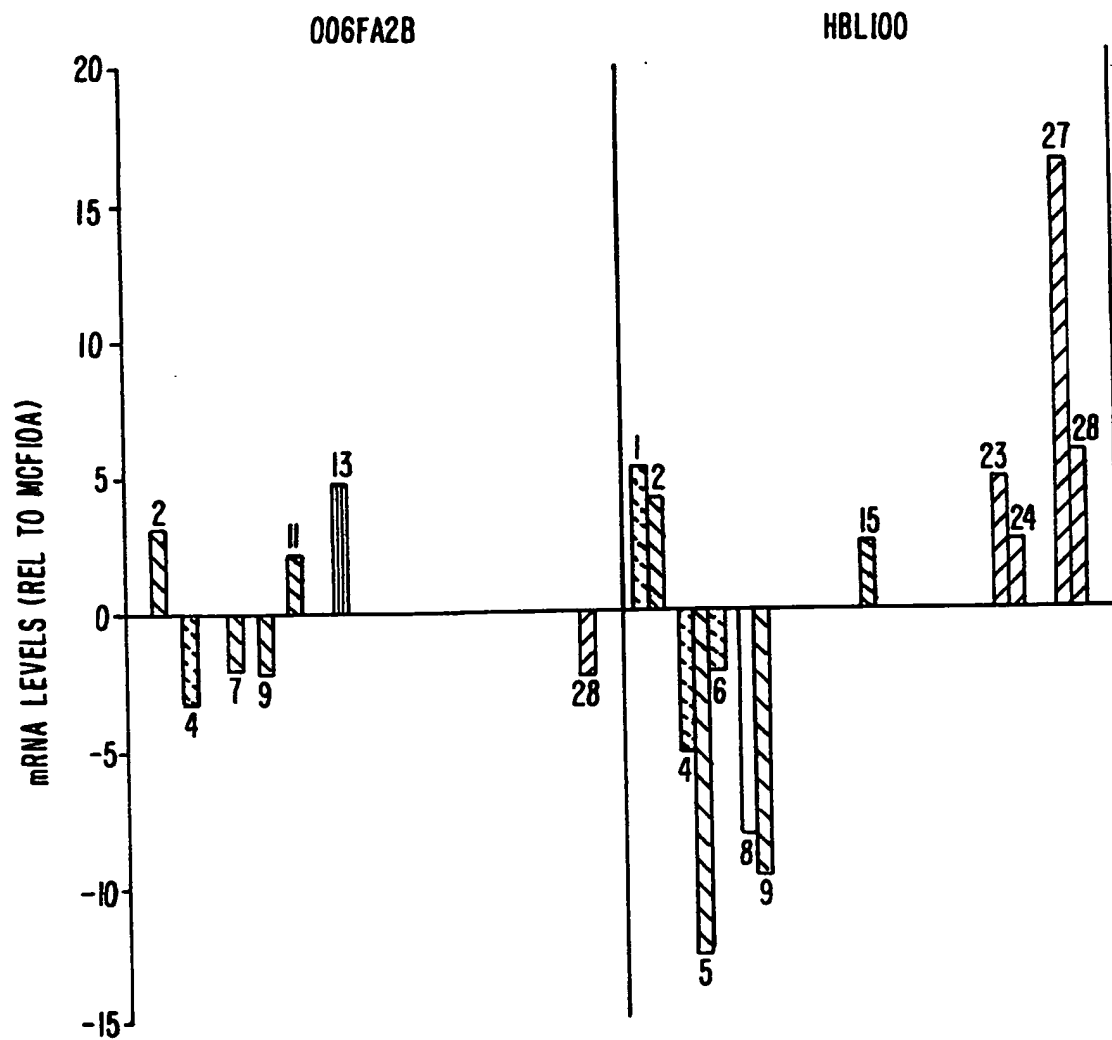


FIG. 3B.

5/6

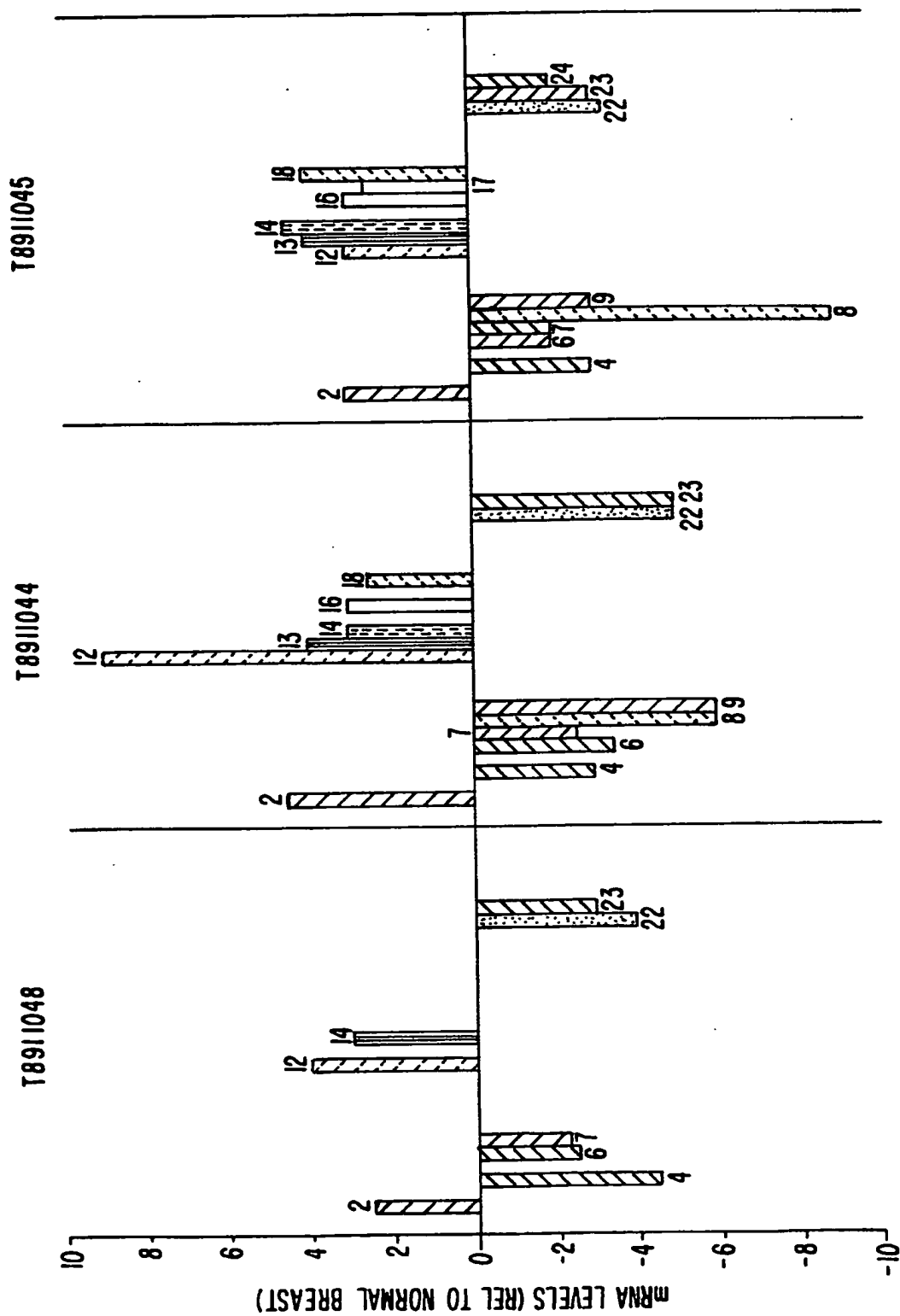


FIG. 3C.

6/6

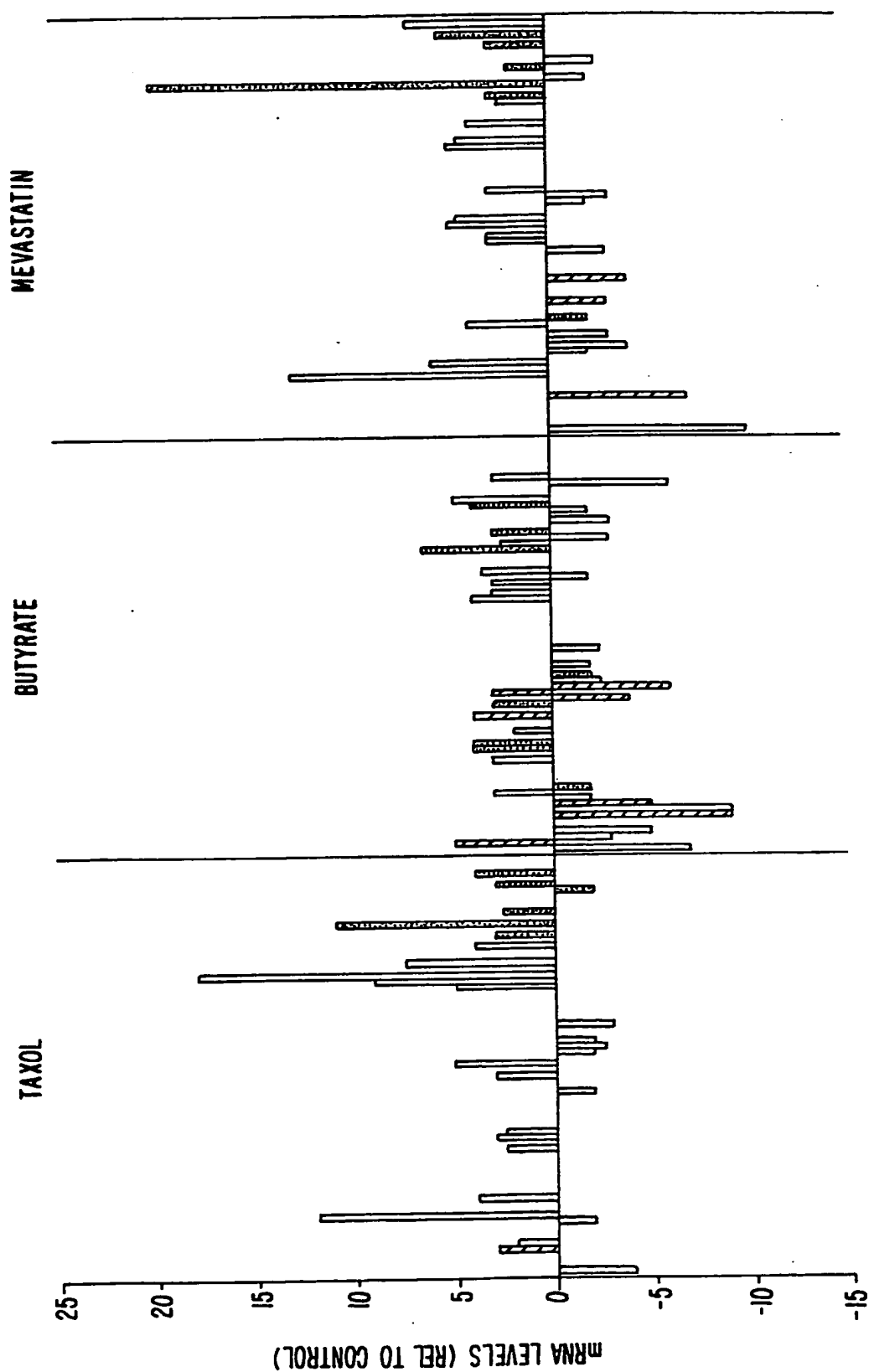


FIG. 4.

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/01552

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| X | WO 97 27317 A (CHEE MARK ;LAI CHAOQIANG (US); LEE DANNY (US); AFFYMETRIX INC (US)) 31 July 1997 see the whole document | 1-59 |
| X | SCHEMA M ET AL: "PARALLEL HUMAN GENOME ANALYSIS: MICROARRAY-BASED EXPRESSION MONITORING OF 1000 GENES" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 93, no. 2, 1 October 1996, pages 10614-10619, XP002070911 See page 10618, third paragraph see the whole document -/- | 1-59 |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "A" document member of the same patent family

Date of the actual completion of the international search

28 May 1999

Date of mailing of the international search report

10/06/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Hagenmaier, S

INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 99/01552

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------|
| A | MAIER ET AL.: "AUTOMATED ARRAY TECHNOLOGIES FOR GENE EXPRESSION PROFILING" DRUG DISCOVERY TODAY, vol. 2, no. 8, 1997, pages 315-324, XP002103832 see the whole document | 1-59 |
| A | HELLER R A ET AL: "DISCOVERY AND ANALYSIS OF INFLAMMATORY DISEASE-RELATED GENES USING cDNA MICROARRAYS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, vol. 72, no. 94, page 2150 2155 XP002076789 see the whole document | 1-59 |
| A | LOCKHART ET AL: "EXPRESSION MONITORING BY HYBRIDIZATION TO HIGH-DENSITY OLIGONUCLEOTIDE ARRAYS" BIO/TECHNOLOGY, vol. 14, December 1996, pages 1675-1680, XP002094438 see the whole document | 1-59 |
| A | FODOR S P A: "MASSIVELY PARALLEL GENOMICS" SCIENCE, vol. 277, 1997, pages 393-395, XP002101908 see the whole document | 1-59 |
| A | SEDLAK B J: "GENE CHIP TECHNOLOGY READY TO IMPACT DIAGNOSTIC MARKETS" GENETIC ENGINEERING NEWS, December 1997, page 1, 14, 34 XP000749581 See "Drug Discovery" see the whole document | 1-59 |
| A | WO 97 10365 A (AFFYMAX TECH NV ; LOCKHART DAVID J (US); BROWN EUGENE L (US); WONG) 20 March 1997 see the whole document | 1-59 |
| P,X | GRAY N S: "EXPLOITING CHEMICAL LIBRARIES, STRUCTURE, AND GENOMICS IN THE SEARCH FOR KINASE INHIBITORS" SCIENCE, vol. 281, July 1998, pages 533-538, XP002101911 see the whole document | 1-59 |

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/01552

| Patent document cited in search report | Publication date | Patent family member(s) | Publication date |
|---|---------------------|--|--|
| WO 9727317 A | 31-07-1997 | AU 2253397 A EP 0880598 A | 20-08-1997 02-12-1998 |
| WO 9710365 A | 20-03-1997 | AU 7073496 A CA 2232047 A EP 0853679 A | 01-04-1997 20-03-1997 22-07-1998 |